J·T·L·A

The Journal of Technology, Learning, and Assessment

Volume 3, Number 7 · February 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

Robert P. Dolan, Tracey E. Hall,
Manju Banerjee, Euljung Chun, &
Nicole Strangman

www.jtla.org

A publication of the Technology and Assessment Study Collaborative
Caroline A. & Peter S. Lynch School of Education, Boston College

JTLA

Volume 3, Number 7

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

Robert P. Dolan, Tracey E. Hall, Manju Banerjee, Euljung Chun, & Nicole Strangman

Editor: Michael Russell
russelmh@bc.edu
Technology and Assessment Study Collaborative
Lynch School of Education, Boston College
Chestnut Hill, MA 02467

Design and Layout: Thomas Hoffmann

JTLA is a free on-line journal, published by the Technology and Assessment Study Collaborative, Caroline A. & Peter S. Lynch School of Education, Boston College.

Preferred citation:

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

Acknowledgment:

This research was supported by funding awarded to CAST from The Peter Jay Sharp Foundation and LD ACCESS Foundation.

J·T·L·A

Preferred citation:

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

Abstract:

Standards-based reform efforts are highly dependent on accurate assessment of all students, including those with disabilities. The accuracy of current large-scale assessments is undermined by construct-irrelevant factors including access barriers, a particular problem for students with disabilities. Testing accommodations such as the read-aloud have led to improvement, but research findings suggest the need for a more flexible, individualized approach to accommodations. The current pilot study applies principles of Universal Design for Learning to the creation of a prototype computer-based test delivery tool that provides students with a flexible, customizable testing environment with the option for read-aloud of test content. Two contrasting methods were used to deliver two equivalent forms of a National Assessment of Educational Progress United States history and civics test to ten high school students with learning disabilities. In a counterbalanced design, students were administered one form via traditional paper-and-pencil (PPT) and the other via a computer-based system with optional text-to-speech (CBT-TTS). Test scores were calculated, and student surveys, structured interviews, field observations, and usage tracking were conducted to derive information about student preferences and patterns of use. Results indicate a significant increase in scores on the CBT-TTS versus PPT administration for questions with reading passages greater than 100 words in length. Qualitative findings also support the effectiveness of CBT-TTS, which students generally preferred over PPT. The results of this pilot study provide preliminary support for the potential benefits and usability of digital technologies in creating universally designed assessments that more fairly and accurately test students with disabilities.

J·T·L·A

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

Robert P. Dolan
Tracey E. Hall
Manju Banerjee
Euljung Chun
Nicole Strangman
CAST

Introduction

Large-scale assessment has become a central component of standards-based reform and a cornerstone of its founding theory of action, which posits that standards, assessment, and accountability work together to improve student learning (Elmore & Rothman, 1999). Crucial to the successful realization of this theory of action are the accuracy and validity of large-scale assessment (AERA, APA, & NCME, 1999). Moreover, in the wake of federal initiatives such as the Individuals with Disabilities Education Act amendments of 1997 (IDEA) and No Child Left Behind Act of 2001, large-scale assessments must be accurate and valid for students within both the general and special education curriculum (Nolet & McLaughlin, 2000).

Unfortunately, available research indicates that the current methods of large-scale assessment are generally inadequate for students with disabilities (Elmore & Rothman, 1999; Hollenbeck, 2002; Olson & Goldstein, 1996; Sireci, Li, & Scarpati, 2003; Thurlow et al., 2000). Particularly problematic is the issue of construct-irrelevance; many assessments measure not only the targeted construct but also unintended constructs related to accessing the test material or carrying out a response (Abedi, Leon, & Mirocha, 2001; Helwig, Rozek-Tedesco, Tindal, Heath, & Almond, 1999; Parkes, Suen, Zimmaro, & Zappe, 1999). Examples of such unintended constructs include sensory capabilities such as sight and hearing, physical capabilities such as holding a pencil, and cognitive capabilities such as memory and attention. These unintended constructs differentially challenge students with disabilities compared to their non-disabled peers.

For students with learning disabilities, who represent approximately half of all students with disabilities, construct-irrelevant difficulty is a pervasive problem. Difficulties with fundamental testing tasks such as selectively attending to a test item and recording responses on a separate answer sheet potentially undermine these students' performance. The reading demands of assessments are a particularly significant source of difficulty for students with learning disabilities, who may struggle with multiple areas of reading literacy: phonemic awareness, phonics/word recognition, vocabulary, fluency, comprehension, and engagement (Ehri, 1994; Graham & Harris, 1996; Helwig et al., 1999; Liberman, Shankweiler, & Liberman, 1989; Stanovich, 1988; Swanson, 1999; Torgesen, 1993). Reading is an unintended construct in many assessments, math generally being one example (Clarkson, 1983; Clements, 1980; Newman, 1977). As such it poses a barrier to some students with learning disabilities, undermining their test performance regardless of their proficiency with the subject or skill area being tested.

Test Accommodations

To address the problem of construct-irrelevant difficulty for students with disabilities, IDEA 1997 requires that students be provided with appropriate test accommodations, alterations in test materials, or procedures to minimize the impact of disability on assessment performance. For students with learning disabilities, the most common test accommodation that alters test presentation is the read-aloud (Sireci et al., 2003; Tindal, Heath, Hollenback, Almond, & Harniss, 1998). The intent of the read-aloud is to level the playing field in terms of reading ability without perturbing in other significant ways the equality of test conditions for students with and without disabilities. While the read-aloud accommodation has been repeatedly shown to improve the performance of test takers with learning disabilities (Calhoon, Fuchs, & Hamlett, 2000; Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000; Meloy, Deville, & Frisbie, 2002; Thompson, Blount, & Thurlow, 2002; Tindal & Fuchs, 1999), there remain issues and concerns with its use.

Most read-aloud accommodations involve a live reading by a teacher or aide to a group of students taking individual tests; less commonly the read-aloud is recorded and presented to a group of students on videotape or audiocassette, and even more rarely students work individually with a computer offering text-to-speech. According to Landau et al. (2003) there are three significant ways in which the human read-aloud accommodations fail to provide adequate supports for students and potentially compromise test validity: 1) human read-alouds vary in quality, and some readers may mispronounce or misread words; 2) students are reluctant to (or may be unable to) ask human readers to re-read test portions (or may be unable to);

and 3) through intonation or non-scripted comments human readers may inadvertently influence students' attention or responses. Human read-alouds, including video and audiotaped read-alouds, also impose upon students a linear navigation path and a set pace, the latter of which has been shown to negatively affect the test performance of students with disabilities (Hollenbeck, Rozek-Tedesco, Tindal, & Glasgow, 2000).

Another shortcoming of the current read-aloud accommodation relates to continuity between instruction and assessment. According to a U.S. Department of Education memo clarifying IDEA with respect to testing accommodations (U.S. Department of Education, 2001), "Assessment accommodations should be chosen on the basis of the individual student's needs and should generally be consistent with the accommodations provided during instruction." Decoding supports for students with learning disabilities vary widely across students and classrooms. During instruction, computer-based text-to-speech tools offer an increasing number of students the benefits of independent, self-paced access to the text (Dawson, Venn, & Gunter, 2000; Farmer, Klein, & Bryson, 1992; Hebert & Murdock, 1994; Lundberg & Olofsson, 1993; McCullough, 1995; Strangman & Dalton, 2005). An equivalent range of options is rarely available to students during large-scale testing. Thus, attempts to simplify and standardize the administration of the read-aloud accommodation are generally at odds with the need to maintain continuity between instruction and testing on an individual student basis.

The homogeneity of the read-aloud accommodation is problematic from another standpoint. The effects of the read-aloud accommodation on different populations of students have not been consistent or easily interpretable (Helwig, Rozek-Tedesco, & Tindal, 2002), with some studies suggesting a differential boost in performance for students with disabilities (Fuchs et al., 2000; Johnson, 2000; Tindal, Heath et al., 1998; Weston, 1999, 2003) and others indicating that these benefits also extend to students without disabilities (Harker & Feldt, 1993; Lee & Tindal, 2000; Meloy et al., 2002). Many scholars have marshaled these findings to question the construct validity of the read-aloud accommodation. Viewed from another standpoint, these patterns of results highlight the need for greater diversity and flexibility in the read-aloud accommodation so that it can be implemented on a more individualized basis. In fact, it has become increasingly clear that not all students with disabilities benefit from the read-aloud accommodation (Helwig et al., 2002; Helwig et al., 1999; Sireci et al., 2003). Responsiveness to the read-aloud accommodation varies by student (Elliot, Kratochwill, & McKevitt, 2001; Tindal, Glasgow, Helwig, Hollenbeck, & Heath, 1998) as well as by other factors such as grade, test form, and problem type (Fuchs et al., 2000; Helwig et al., 1999), and researchers have called for greater attention to individual effects of accommodations (Elliot et al., 2001; Helwig et al., 1999; Tindal, Heath et al., 1998).

Universal Design and Assessment

The limited efficacy and unintended effects of accommodations such as the read-aloud have led to the consideration of an novel (Sireci et al., 2003). direction for assessment that is embodied by two related approaches, Universal Design for Learning (UDL) (Orkwis & McLane, 1998; Rose, Meyer, Rappolt, & Strangman, 2002) and universal design of assessment (Thompson, Johnstone, & Thurlow, 2002). Both of these approaches build on the universal design movement in architecture, where architects seek to avoid costly and inefficient retrofits by designing buildings with the needs of all potential users in mind. Universal design of assessment seeks to apply the universal design principles (Connell et al., 1997) originated by Ron Mace to the design of tests compatible with broad student participation (Thompson, Johnstone et al., 2002). Preliminary research findings suggest that students with disabilities, indeed all students, may perform significantly better on tests applying universal design principles than on traditionally designed tests (Johnstone, 2003). UDL is a theoretical framework grounded in the notion that curriculum designers can increase the likelihood that all students will be able to successfully access both content and learning in the curriculum by considering from the start the diverse ways in which these students learn (Orkwis & McLane, 1998; Rose et al., 2002). The three UDL principles guide the design of goals, methods, assessments, and materials that together accommodate student diversity by providing students multiple, flexible opportunities for recognizing information, strategically interacting with the curriculum, and engaging with the curriculum. Specific considerations for applying UDL to large-scale assessments have been proposed (Dolan & Rose, 2000; Dolan & Hall, 2001). While UDL conceptually applies to traditional media and instructional approaches, technology is seen as a key enabler due to its inherent flexibility, which makes an individualized approach more feasible.

Advances in technology have made computer-administered testing a possibility. While the overall focus on computer use has been on decreasing costs and increasing timeliness associated with large-scale testing, a few states , most notably Kentucky (HumRRO, 2003; Salyers, 2002), have been exploring the role of technology in improving statewide test accessibility for students with disabilities. In addition, guidelines are emerging for making computerized tests most accessible (Allan, Bulla, & Goodman, 2003; Association of Test Publishers, 2002; Thompson, Thurlow, Quenemoen, & Lehr, 2002). A computerized read-aloud is a potentially valuable tool for realizing a universal design approach toward assessment and addressing the current problems inherent within the human-mediated read-aloud accommodation. Unlike a human read-aloud, a text-to-speech read-aloud provides students with consistent readings free of potentially directive or misleading intonation. The use of text-to-speech can increase continuity between instruction and assessment, by broadening the range of options available during testing. And consistent with UDL, a text-to-speech read-aloud supports students' diverse ways of recognizing, strategically interacting, and engaging with an assessment by offering individualized, independent, and self-paced multimodal access to test content – on demand.

Four recent research studies (Brown & Augustine, 2000; Burk, 1999; Calhoon et al., 2000; Hollenbeck et al., 2000) support the effectiveness of a computerized read-aloud during testing. Studying the effectiveness of screen reading software for high school seniors with and without reading difficulties taking the National Assessment of Educational Progress (NAEP) science and social studies tests, Brown and Augustine (2000) concluded that the technology could be useful for poor readers. Calhoon et al. (2000) reported that accommodation in the form of a computer read-aloud, with or without video, significantly increased scores of students (mean age 16 years) with learning disabilities on a math assessment. Hollenbeck, Rozek-Tedesco, Tindal & Glasgow (2000) looked specifically at the benefits of self-pacing in a computer read-aloud and demonstrated that students with disabilities performed better on a math test when using the self-paced, computer read-aloud versus a video read-aloud. Finally, Burk (1999) found that high school students and adults with learning disabilities scored significantly higher on computer-based tests that provided large print, extra spacing, and recorded human voice compared to standard paper-based delivery.

Current Study

The pilot study presented here investigates the potential of computer-based read-aloud testing accommodations. The performance of students with learning disabilities on a multiple-choice United States history and civics test was compared under two conditions: one using a computer-based testing system with text-to-speech read-aloud capability and one using a traditional paper and pencil version of the test. Unlike previous studies of computerized read-alouds, this study investigated not only group-wide effects but also the impact of the accommodation on individual students. As such, the study addressed the potential importance and effectiveness of more flexible and individualized assessment built on the principles of Universal Design for Learning.

Research Questions

The overarching research question addressed in this study is whether computer-based testing with text-to-speech (CBT-TTS) is an effective approach for providing individualized support to students with learning disabilities (LD) during multiple-choice testing. From this are derived the following research questions:

Is CBT-TTS effective?

Is CBT-TTS an effective means for assessing high school students with LD compared to traditional paper-and-pencil tests (PPT) with no read-aloud accommodation, as measured by changes in test scores and student impressions and preferences?

What aspects of CBT-TTS make it effective?

To the extent that CBT-TTS is an effective means for assessing high school students with LD, which components of the system may be responsible?

Would students use CBT-TTS in the real-world?

Would students choose to use CBT-TTS during testing if it was available?

Methods

Participants

Prospective participants with specific learning disabilities were identified by soliciting recommendations from resource room teachers at a suburban public high school. Fifteen 11th and 12th grade students were recommended and volunteered to participate in the study. All were served in special education with an active Individualized Education Programs (IEP) and were partially or fully included in general education classes. Five students were later excluded because they did not have a diagnosis of LD in their IEPs; an additional student with comorbid emotional disturbances was excluded based on an inability to follow directions.

Procedure

The study took place over a three-week period. Students were informed that the study was evaluating new test administration procedures involving computers. Each student was administered two equivalent forms of a U.S. history and civics test on two separate days. One form of the test was administered using traditional paper-and-pencil testing (PPT) methods, while the other form was administered using CBT-TTS. To control for order effects and differences between the test forms, order of administration (PPT first vs. CBT-TTS first) and test form (A vs. B) were counterbalanced across 4 randomly composed groups. Students were assured that while they would be expected to do their best, test scores would not affect their grades.

Before taking the first test, a member of the research team trained on use of the CBT-TTS system. During this training, which occurred in a group setting, students' computer interactions were observed and it was verified that each student possessed the requisite skills to use the system.

For the CBT-TTS administration, students were seated separately with individual laptop computers. In the event that students were not familiar with the pointing devices built into the laptop computers, standard mice were provided. Students were also provided with headphones to ensure that they would not be distracted by each others' use of text-to-speech supports. All computers ran the Microsoft Windows 2000™ operating system. Students were not provided with a paper test booklet. For the PPT administration, students were seated separately with their test booklets and a pencil. Both the PPT and CBT-TTS administrations were conducted by a member of the research team, who also read administration instructions to the students.

Students were allowed up to 45 minutes to complete the test. This was based on guidelines provided by NAEP staff (Lazer, 2002, personal communication).

The training session occurred during school hours in the school library. Testing and interview sessions took place after school in a resource room. All sessions lasted approximately 50 minutes.

Materials

Assessment Instrument

The two test forms were assembled using released items from the NAEP U.S. history and civics tests. To confirm that reading ability was an unintended construct for these test items, such that they would remain valid with use of a read-aloud accommodation, content classification data provided on the NAEP website (National Center for Education Statistics, 2002) were evaluated. Twenty-two passages were selected, each one was accompanied by one or two multiple-choice items. As is described in greater detail below, the selected passages and accompanying items were matched and then divided into two test forms. Each test form contained a total of 15 questions across the 11 item sets. Discussions with the high school's history teacher indicated that students had been exposed to the content covered by the tests.

The two forms were matched in terms of item difficulty, content area, cognitive domain, reading passage length, and readability of stimuli and questions. Item difficulty, content area, and cognitive domain were determined from the NAEP Questions Tool (National Center for Education Statistics, 2002). Reading passage lengths were calculated electronically using word count software. Readability of stimuli and questions were determined using the Lexile Analyzer software (MetaMetrics, 2004; White, 2001), which rates text according to the Lexile Framework using measurements of word frequency and sentence length.

Both the PPT and CBT-TTS test administrations included two accommodations typically provided for students with LD. The first accommodation was presentation of test item sets (i.e. a reading passage and associated question or questions) one at a time, which allowed students to focus on the test item sets without being distracted by others. The second accommodation was elimination of a separate answer sheet, which allowed students to respond directly on the test booklet or computer-screen, without having to locate and transpose their responses. In addition, the CBT-TTS administration included a third accommodation, text-to-speech-based read-aloud. The read-aloud accommodation provided students with a spoken representation of test passages, questions, and responses, supporting any decoding challenges they might have and thus allowing them to dedicate their efforts to the intended subject-area-centric constructs.

Test Delivery System

A prototype computer-based testing system was designed to deliver the CBT-TTS condition evaluated in this study. Within this system, one at a time presentation of item sets was provided by displaying only related test questions and passages at any one time. "Clickable" radio buttons next to each multiple-choice response were used in place of a separate answer sheet. The CBT-TTS system was developed using accessible HTML (Hypertext Markup Language) (W3C-WAI, 1999) so that it could be viewed using a "talking browser" in an accessible manner. CAST *eReader*™ software was used to provide this text-to-speech support, allowing students to have select words, sentences, or entire passages read and reread individually at will. Read-aloud support was available for reading passages, test questions, and responses. CAST *eReader*™ software also allowed students the option of viewing synchronized highlighting of words, which provided visual feedback that facilitated independent reading and navigation, as well as decoding. CAST eReaderTM software was one of several readily available TTS software applications that provide the same general functionality.

An additional design emphasis of the CBT-TTS system was to provide students with choices in how they proceed through the test. Students have many options when taking traditionally-administered tests, such as the order in which they answer questions, the ability to review items or look ahead, and the ability to read and reread passages, questions, and responses in an arbitrary order. The CBT-TTS system was designed to offer similar flexibility, while still being easy to learn and use. As can be seen in the sample screenshot from the CBT-TTS system in Figure 1, students can view simultaneously the *reading passage* and the related multiple-choice *questions and responses*. Students can optionally mark individual questions that they wish to review later. Also, a *navigation bar* allows students to view their progress through the test, see which items they have completed and/or marked for subsequent review, and jump to items in any order.

(Figure 1 is shown on the following page.)

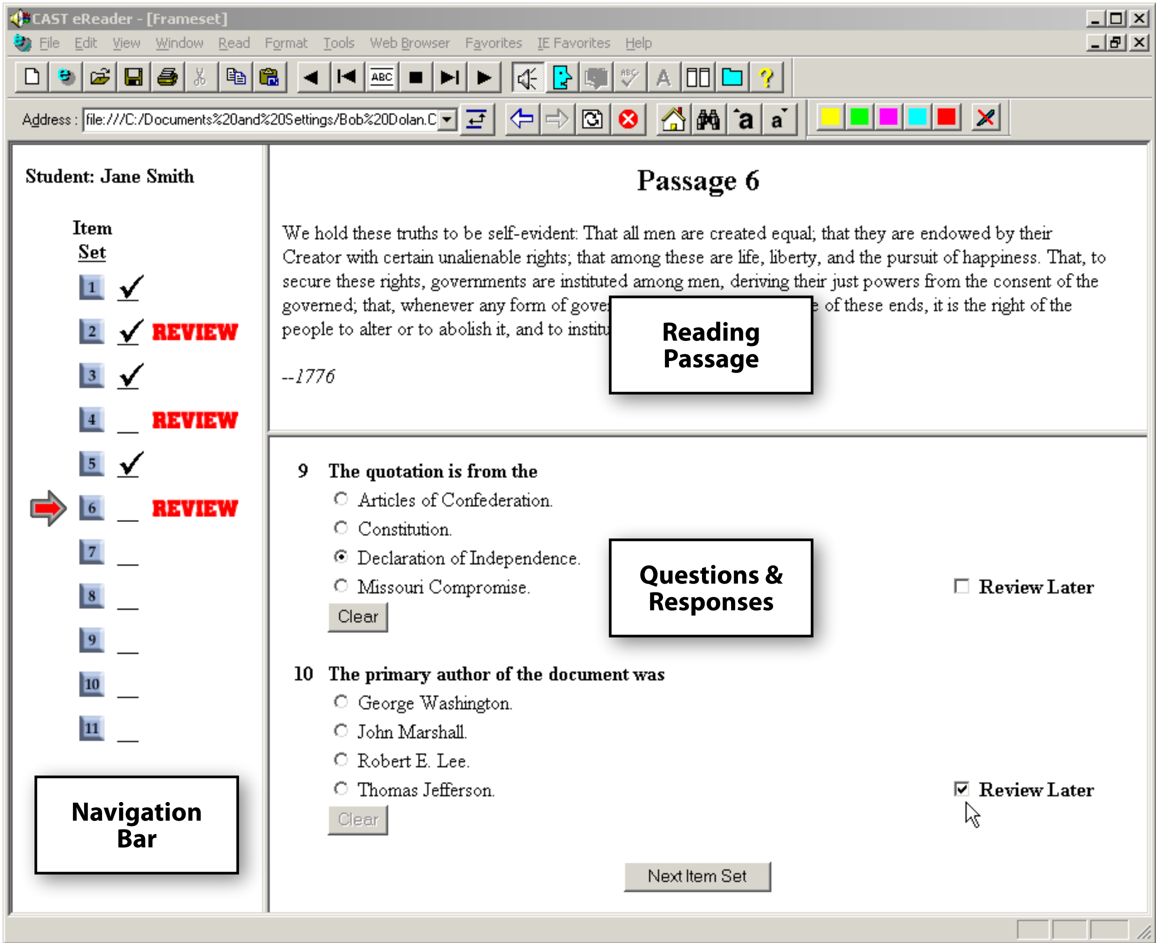

Figure 1. Screenshot of the prototype computer-based testing system with text-to-speech support (CBT-TTS).

Data Collection

Independent Variables

Measures of students' reading abilities were collected through individual administration of the WIAT®-II reading subtests: pseudoword decoding, word reading, and reading comprehension. A reading composite score was calculated for each student based on their subtest scores.

Dependent Variables

Student test scores were separately calculated for each administration condition and used to compare the efficacy of CBT-TTS and PPT. Four additional data sources were used to investigate students' patterns of use and impressions of the system as well as pre-existing opinions of test-taking, accommodations, and computers. These additional data sources include: usage tracking, field observations, student surveys, and structured interviews. During administration of the CBT-TTS, usage-tracking technology recorded students' progress and the frequency with which they used the various system features. CAST researchers performed observations of participant engagement and behavior during both administrations. After finishing both tests, students completed a thirty-item survey designed to capture their opinions about prior experiences with computers, test-taking, and accommodations, strategies that they use during test taking, and their impressions of the CBT-TTS system. In addition, six students were interviewed to obtain further impressions, suggestions, and feedback; these students were chosen based on teacher recommendations as being comfortable with an interview and likely to be talkative.

Data Analysis

Test Scores

To evaluate the effects of administration condition on overall student test performance, mean test scores for PPT and CBT-TTS were compared statistically. As a part of this analysis, scores on subgroups of test questions were analyzed as a function of the length of the stimulus reading passage. A matched sample comparison of means (t-test) was conducted across three different sets of test stimuli: all reading passages, long passages (more than 100 words), and short passages (100 words or less). A correlation statistic 'r' based on a transformation of the Cohen's *'d'* index (Sheskin, 2000) was used to measure effect size.

Student Preferences and Usage of Test-taking Strategies

Student surveys, structured interviews, field observations, and usage tracking data were qualitatively analyzed to identify students' actual test-taking behavior and stated preferences about test-taking. Table 1 describes the instruments used to provide the qualitative information.

(Table 1 is shown on the following page.)

Table 1: Instruments Used to Provide Qualitative Information on Students' Test-Taking Behaviors And CBT-TTS Feature Preferences

Test-taking behaviors and feature preferences	Survey	Interview	Field Observation	Usage Tracking
Overall Impressions of CBT				
Perceived impact on performance	X			
Access to test	X			
Comparison of CBT-TTS and PPT				
Usefulness of CBT-TTS vs. PPT		X		
Preference for CBT-TTS vs. PPT	X			
CBT-TTS Usability				
Clarity of directions	X			
Technical difficulties	X			
Use of CBT-TTS Features				
Navigator bar	X		X	
Return to previous question			X	X
TTS voice	X	X		
Response changes			X	X
Scrolling to read passages/questions	X			
Use of CBT-TTS Test-taking Strategy Supports				
Review Later marker	X		X	X
Linear/non-linear progression			X	X
Viewing of completed item set check marks			X	X
Highlight words while reading/on test		X		X
Use of CBT-TTS Read-aloud Support				
Decision when to use TTS		X		
TTS to read passages	X			
TTS to read questions	X	X		
TTS to read test responses	X			
TTS to aid comprehension of passages	X			
TTS to aid understanding of test questions	X			
TTS to re-read	X			
CBT-TTS Test Item Display Format				
One-at-a-time item set presentation		X		
Perceived impact of presentation on performance		X		
CBT-TTS Test Response Formats				
Elimination of separate answer sheet		X		
Screen response vs. bubble sheet	X			

An "X" indicates that an instrument provided information on a given behavior or preference.

Results

Test Performance with PPT vs. CBT-TTS

As seen in Figure 2, students performed slightly better on the CBT-TTS version of the tests than on the PPT version. Specifically, students answered 65.3% of the items correctly when performing the test with the CBT-TTS versus 58.7% with the PPT version. This score difference represents an effect size of 0.49, but is not statistically significant ($t = 1.71$; $p = 0.12$). However, as also seen in Figure 2, the pattern of performance differed between long and short passages. When responding to items associated with long reading passages, students scored approximately 22 percentage points higher on the CBT-TTS administration (mean percentage score 76.7%) than the PPT administration (mean percentage score 55%). This score difference represents an effect size of 0.6 and is statistically significant ($t = 2.26$; $p = 0.05$). In contrast, students performed slightly better on the PPT version (mean = 60.0%) than the CBT-TTS version (mean = 58.0%) when responding to items associated with short passages. This score difference, however, represented an effect size of only 0.29 and is not statistically significant ($t = 0.90$; $p = 0.39$).

(Figure 2 is shown on the following page.)

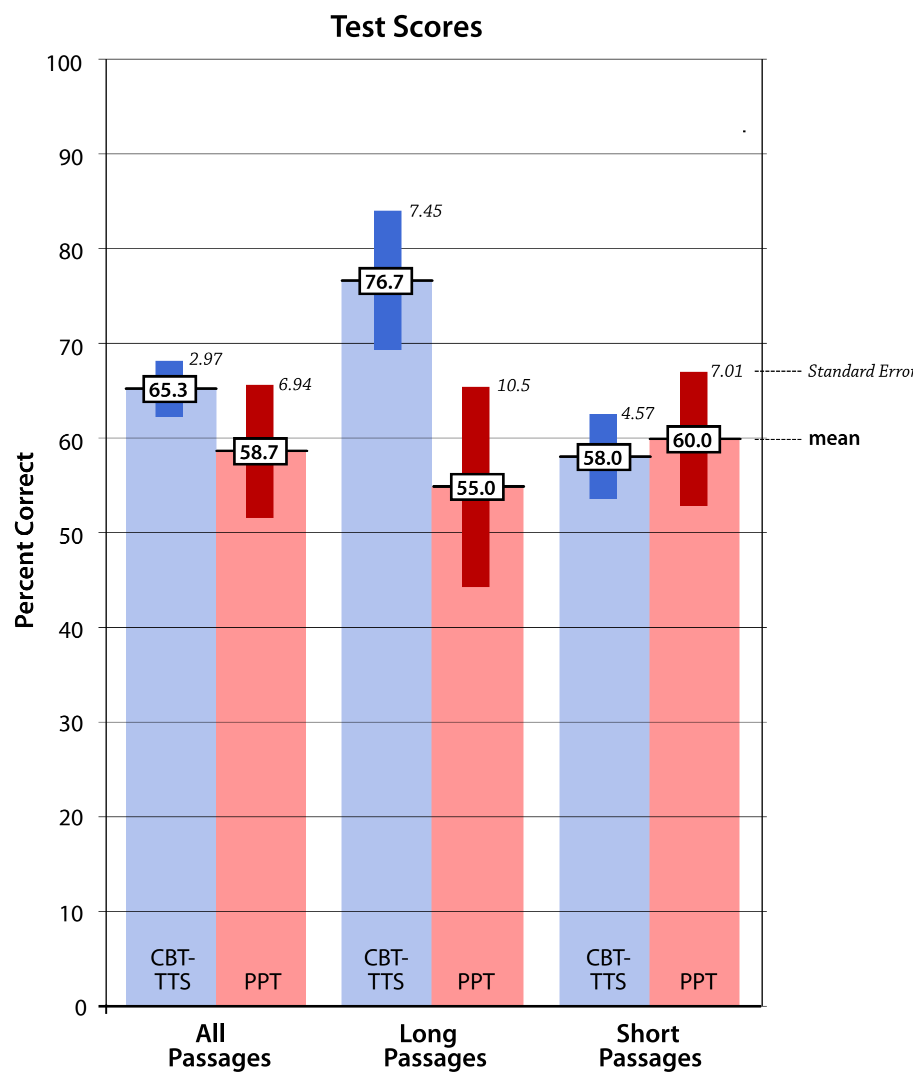

Figure 3. Mean test scores for CBT-TTS versus PPT administration for all test questions, test questions with long passages, and test questions with short passages.

To further investigate the improvement in test scores for the long passage questions, the relationship between students' performance with the two test versions was examined. For these analyses, students' WIAT®-II reading composite score were used to categorize students as "low average" readers (WIAT®-II reading composite scores below 80) or "average" readers (WIAT®-II reading composite score above 80). Of the seven students who performed higher on the CBT-TTS version as compared to the PPT, three were classified as "low average" readers. None of the three students who performed lower on the CBT-TTS version than the PPT version were classified as "low average" readers. Thus, not only did noticeably more students perform better on the items associated with long passage items when using the CBT-TTS version, all of the students classified as "low average" readers performed better on the long passages when using the CBT-TTS version. Specifically, these students performed 17, 42 and 75 percentage points higher on the CBT-TTS version.

In contrast, when examining the short passage items, half of the students performed better on the CBT-TTS version and half performed better on the PPT version. The three students classified as "low average" readers, however, were split between these two groups, one performing 14 percentage points higher on the PPT version and the others performing 23 and 10 percentage points higher on the CBT-TTS version. Due to small sample size, it was not possible to perform a regression analyses to further explore the relationship between reading skills and performance under the different conditions.

Form and Order Effects

No statistically significant difference was noted in CBT-TTS test scores (p = 1.0) between Forms A (mean score = 65.3%) and B (mean score = 65.3%). Similarly no statistically significant difference was found in PPT test scores (p = 0.58) between Forms A (mean score = 54.7%) and B (mean score = 62.7%).

No statistically significant difference was noted in CBT-TTS test scores (p = 0.76) when PPT was administered before CBT-TTS (mean score = 64.0%) versus when PPT was administered after CBT-TTS (mean score = 66.7%). PPT test scores differed, but not significantly (p = 0.35), when PPT was administered before CBT-TTS (mean score = 65.3%; p = 0.35) versus when PPT was administered after CBT-TTS (mean score = 52.0%), evidence for a trend toward better performance when the PPT version was taken first.

Student Interaction With and Perceptions of Testing Environments

Overall, student impressions of the CBT-TTS were uniformly positive. An overwhelming majority of students (90%) reported having very few, if any, technical problems in using the CBT-TTS system; all students reported that they easily understood how to use and navigate within the computer-based environment. Students said that the CBT-TTS was "easier to use" and "easier to understand" than the PPT test. Further analysis of the qualitative data suggests that student preferences for CBT-TTS are linked to features that promote independence and flexibility for commonly used test-taking strategies. In addition, students strongly endorsed the text-to-speech feature within the CBT-TTS format.

Qualitative data on students' fluency and past familiarity with computers were analyzed to investigate potential relationships with preferences for test format and/or test features. A majority of students (70%) self-reported their ability to use computers as either excellent or good; 80% of students rated their comfort level in using computers for tests or schoolwork as excellent or good. Reporting on familiarity with computers, 70% of students said that they use the computer frequently at home; 90% reported a high level of familiarity with web browsing; and 80% of students said that they use word processing capabilities a lot/often on the computer. No relationship was found between students' self-reported computer skills and their CBT-TTS preferences or usage. However, students who indicated their computer skills to be less than excellent or good did self-report low usage of certain features on the CBT-TTS in this study, such as the "Review Later" marker.

Although most students (70%) reported having no or limited prior experience with TTS on a computer, survey data showed that given the option, a large majority of them (90%) used this feature often to read the test question passages in the study. This finding is further supported by field observations, which confirm use of TTS by all students to read passage text. In an interview one student said, *"I figured I should use it [text-to-speech] for the experience and it was easy - better than having to read alone and it helps to comprehend."* Another student reported, *"I used the text and speech for all [passage, question, and answer reading]. It made it easier and read directly to the brain. I made it keep reading until I got it. It kept me from boring [being bored]."* Interview data suggests some students preferred to read along with TTS from the beginning, while others used it after initially reading the text by themselves.

Reponses were mixed regarding the use of TTS for reading the question sentences and the answer options on the multiple-choice test format; 40% of students said they used TTS a lot to read the questions and answers, while another 40% said they used TTS only sometimes for this purpose; 20% of students said they rarely used TTS to read the question sentences and answer options.

As an accommodation for reading disabilities, the use of TTS as compensatory support differed among students. Ninety percent of students noted using TTS to read (decode), and 70% said TTS definitely helped their comprehension of the passages. When asked if they used the TTS to help understand the question and answer options, 20% of students responded they used it a lot to help them comprehend, and 50% said they used it frequently though not a lot to aid their comprehension of question and answer options. As to how they decided when to use TTS, one interviewed student replied, "*[I] used it on passages that had so much information*." In general, students indicated a preference for TTS over a human reader. Several interview statements were particularly revealing on this topic. When asked about their preference for receiving the read-aloud accommodation using TTS on a computer versus a human reader, one student said, *"Text-to-speech is easier to use. You can see and at the same time listen to it. I have more control when I use the computer."* Another student stated, *"It is hands-on for me. More control on what gets to me, and I can tell myself. It was easier for me because I am getting it."*

We collected data on student usage and preferences for two other comparable features between the CBT-TTS and PPT formats, namely, test item display and item navigation. Students clearly liked having the test passage, test question, and answer options all on the same screen. One student said, "*When I saw the answers, I could go back and match it with key terms or sentences (in the question) and click the answer*." When asked about taking tests administered traditionally using PPT, all students reported a preference for being able to respond without the use of a separate answer sheet. One student captured these sentiments well during an interview: *"Bubble sheets are horrendous. I get confused with them too. There is too much on one page - they make me nervous. I lose my place and can't keep track, so I can't concentrate on the test itself. I'm worried about marking in the right place*."

Survey data revealed that 70% of students looked at the navigation bar a lot/frequently during the CBT-TTS administration, and an equal number of students (70%) reported finding it to be very useful or useful. Students interviewed confirmed that they use markings in the test booklet as a "*mind hanger*", and as a feature that helps them to remember what to do on the test. CBT-TTS usage tracking data indicated that 40% of students actually used the navigation tools to mark test items. However, usage of the "Review Later" feature was varied. While 40% of students said they never actually used the review marker feature, another 40% said they used it sometimes. Students with LD commonly resort to physical tracking of words and sentences while reading. In this study, 60% of students were observed following along with their finger or pencil while reading test items on the PPT, and 100% of students were observed either using the cursor or their finger to track on the screen while reading.

A majority of the students (60%) thought they did better on the CBT-TTS administration of the test, and nearly all students responded with a "definite yes" for recommending the use of CBT-TTS for other students.

When asked if they had any suggestions for changes to the CBT-TTS system, interviewed students provided the following insightful comments:

- *"I like it that you can just read a word, read more or the whole thing. It's good to make it be able to go faster, slower and change voices."*
- *"[Found problems with] reading hyphenated words. Oh, reading with expression, read with tone that is more moderate in pitch and regulated. It helped, but it's not as good as it could be."*
- *"I think you have to work on voices in eReader. It would be better if I can use color-coded tests and shading as in MCAS [Massachusetts Comprehensive Assessment System] or SAT. It was much better that I can see and hear at the same time and also can play with it."*
- *"Use pictures; likes visuals, such as an image of Thomas Jefferson signing the Declaration of Independence rather than just his name.*
- *Mary's voice got annoying over time. Mispronunciations and accenting (e.g. Martin Luther King) were annoying."*
- *"Voices. I chose Mary because it was more humanlike. Echoes were obnoxious. I think it would be better to use it when I read books not just for taking a test."*

Discussion

The results from this study indicate that providing computer-based read-aloud support to high school students with learning disabilities can improve their performance on a multiple-choice United States history and civics test. Quantitatively, we found a large and statistically significant increase in scores on the CBT-TTS versus PPT administration for questions with reading passages greater than 100 words (roughly one paragraph) in length.

Our qualitative findings also support the effectiveness of CBT-TTS, with students generally preferring this form of administration over PPT. Students further indicated that the TTS features of the CBT-TTS test administration were the most helpful; while they appreciated the navigation and accessibility features, there is little indication that these alone would have made much difference to them. Their usage patterns and responses also provide strong support for this approach being used in real-world situations. The possibility of a "novelty effect" contributing to students' favoring of CBT-TTS cannot be ruled out. However, a strong novelty effect would not predict the differences in students' use and ratings of the CBT-TTS features.

The lack of significant difference in test scores between the two accommodation conditions for questions with shorter passages is not entirely unexpected. For one, readers are likely to struggle less while reading shorter passages, thus reducing the benefits of TTS support; our sample size may have been insufficient to detect this smaller difference in scores. It is also possible that for some students the computer created new accessibility barriers, such as an unfamiliar interface or having to read on a computer screen, that could only be overcome with the advantages of read-aloud for long passages. In a more general sense the conditions of this study were not optimized to realize the full potential benefits of CBT-TTS. Students lacked extensive experience with the system. Also, students were informed that this test would not in any way affect their grades, a fact that may have led them to exert less effort than they would have in a live testing situation. With more extensive training and familiarity with the system, and in a live testing situation, students might use the TTS feature more effectively and to their greater benefit – on both short and long passages. While we consider the current study successful in addressing our research questions and demonstrating the potential utility of computer-based test read-aloud administration, replication with a larger sample size drawing from more than one school is needed to examine whether the findings generalize. It is also important to investigate the impact of student age, race, socioeconomic status, computer experience, and differences in content area knowledge on the efficacy of CBT-TTS.

Our findings generally agree with those of others who have found computerized read-aloud during testing effective for students with disabilities (Brown & Augustine, 2000; Burk, 1999; Calhoon et al., 2000; Hollenbeck et al., 2000). In addition, our qualitative findings corroborate teacher and student impressions uncovered during Kentucky's initial implementation of their online assessment (HumRRO, 2003). For example, one teacher who proctored the Kentucky online test stated she saw "*students re-read questions when they would likely not have asked a human reader to repeat the item.*" Another mentioned that "*the computer version appeared to keep student's attention, particularly for longer reading and math items*." One student who took the online test for the first time said that when he saw long questions last year, he guessed at the answers, but this time he understood the questions better and tried to do well. Based on these observations, a direct comparison of CBT-TTS and traditional administration with human readers is recommended as a future study.

It should be noted that our order effect analysis uncovered a decrease in PPT scores for students that took the PPT after the CBT-TTS. While not a significant difference, it is possible that the engagement of these students was diminished while taking the test the second time without the novelty of the CBT-TTS system. Only with larger sample sizes could such an effect be properly evaluated.

A dominant concern in both the assessment research and the testing communities is whether the read-aloud accommodation differentially benefits students with and without disabilities. The current study does not address the interaction hypothesis (Shepard, Taylor, & Betebenner, 1998; Sireci et al., 2003; Zuriff, 2000). However, we must consider the following: (1) the intended constructs for the NAEP U.S. history and civics items do not include decoding ability, and (2) students diagnosed as having LD are not the only ones who could benefit from decoding supports. Thus while evaluating the interaction hypothesis is important from a research perspective, from a pedagogical perspective it may artificially constrain the utility of CBT-TTS as means for ensuring test fairness and accuracy. While test comparability is of paramount concern in large-scale assessment, current notions of comparability may be based more on limitations in current psychometric techniques than on a pedagogical understanding of learning and the demonstration of learning.

The motivation behind the current study was not only to improve existing accommodations, but to explore how the principles of universal design might apply to the delivery of large-scale assessments. As such, we prototyped and evaluated a test delivery tool which allows students flexible means for demonstrating their knowledge and skills, without compromising test validity. Universal Design for Learning in particular is based on the belief that one size rarely fits all. For testing, this implies that students with equivalent construct-relevant knowledge and abilities may perform differently during a standardized test administration simply because of construct-irrelevant differences Unfortunately, many testing accommodations, while accommodating for some sources of construct-irrelevant difficulty, still take a one-size-fits-all approach to supporting students. The use of human readers, for example, can compromise some students' ability to self-pace, a situation further aggravated during group administration, and thus negatively affect test performance. The CBT-TTS prototype evaluated in the current study fulfills UDL's recommendation that students be provided multiple, flexible means of representation of information, namely text and audio, with opportunities for simultaneous presentation if possible, namely synchronized highlighting. The system also allows students to proceed through the test in any order, to read questions before passages, and the like, as well as adjust font size and voice parameters. As such, the system supports individual student's differences, not a generalized model of student.

While improvements to assessment delivery systems can go a long way in creating tests that reduce the effects of construct-irrelevant factors, they can only go so far. The approach to universal design must begin during test development (Dolan & Hall, 2001; Thompson, Johnstone et al., 2002). For example, assumptions are still often made about students' cultural experiences and reading comprehension abilities (on subject area tests); these assumptions cannot be "accommodated away" through retrofitting. Instead, they must be addressed during item development. Equally important is that the use of technologies such as TTS be matched with students' abilities and challenges as readers, and that such matching start in the classroom. The fact that few students in the study had prior experience with TTS tools underscores the need for more appropriate use of such accommodations during teaching, not only during testing. Only then can we be assured that students are receiving the supports necessary to ensure their ability to learn, and their ability to be assessed fairly and accurately.

Based upon student comments, certain modifications to the CBT-TTS system may be useful. Most importantly, students found limitations with the TTS technology, namely mispronunciations and strange prosodies. We expect ongoing research and development on synthesized voices to have a dramatic effect on this technology. Additional new technologies, such as the embedding of recorded human voices using DAISY-3 technology (ANSI/NISO Z39.86-2002), are under investigation by several researchers and might prove suitable for use during testing in addition to or as an alternative to TTS. In addition, the tagging of text with speech generation directives using schemes such as the Speech Synthesis Markup Language (WAI, 2004) may improve the quality and consistency of TTS in the future.

Delivery of large-scale assessments using TTS technologies is currently being implemented in only a handful of locations nationwide, but it is reasonable to expect much wider spread implementation over the next few years. Fortunately, the technologies employed in this study have become largely ubiquitous. A number of TTS software applications are currently available, some even free, for both Windows™ and Macintosh™ operating systems. Furthermore, with the recent endorsement of the National Instructional Materials Accessibility Standard (NIMAS) by the U.S. Department of Education (U.S. Department of Education, 2004), students will have ever increasing opportunity to interact with digital instructional materials.

A significant challenge lies in determining just how well matched the instructional and testing technologies must be. For example, in implementing the CATS Online, the Kentucky Department of Education ensured that all students had access to exactly the same TTS software they used in the classrooms. For commercial test publishers developing solutions for use in multiple states, however, this requirement may prove problematic. It is therefore imperative that we better understand just how much difference in implementation is tolerable to students and is still "consistent with the accommodations provided during instruction" (U.S. Department of Education, 2001).

We recommend that the current pilot study be followed up with additional larger-scale studies to continue this line of research. Additional accommodations, especially ones that support alternate means of response such as use of computer-based writing supports, would allow students independence while completing short answer and open response questions. Furthermore, a better understanding of the effect of training is critical to ensuring students properly master the testing environment. This is especially important as new testing technologies emerge in testing before they have had time to be fully integrated into instruction. Finally, the use of technologies such as TTS in other subject areas, especially math and science, remains less understood, since much less is known about how these supports best work during instruction.

Conclusion

The current study expands the small but growing body of evidence implicating the use of digital technologies in creating universally designed assessments that more fairly and accurately test students with disabilities. Even before tests are created fully in accordance with universal design principles, and even with the known limitations of the technologies, the results of this study have implications that can be applied today by educators and test publishers as they explore ways to increase not only the quantity of testing, but also the quality of the results. This will help ensure that students with disabilities, and eventually all students, are tested on their construct-relevant knowledge and skills rather than on construct-irrelevant factors.

It is important to remember that the goal of universal design is to support all users, not only those with disabilities. As such, any testing solutions that reduce construct irrelevancy will improve the validity of decisions made upon test scores. To this extent we must be willing to embrace assessment techniques that provide students with the best opportunity to demonstrate their knowledge and skills, even at the expense of presentation "consistency"; in fact, consistency has been little more than illusion given the extreme diversity in the ways in which individual students develop and demonstrate knowledge and skills.

References

Abedi, J., Leon, S., & Mirocha, J. (2001). *Validity of standardized achievement tests for English Language Learners.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, Washington.

AERA, APA, & NCME. (1999). *Standards for Educational and Psychological Testing*. New York: AERA.

Allan, J. M., Bulla, N., & Goodman, S. A. (2003). *TEST ACCESS: Guidelines for Computer Administered Testing*. Louisville, Kentucky: American Printing House for the Blind.

Association of Test Publishers. (2002). *Guidelines for Computer-Based Testing*. Washington, DC.

Brown, P. B., & Augustine, A. (2000). *Findings of the 1999-2000 screen reading field test*. Delaware Department of Education.

Burk, M. (1999). *Computerized test accommodations: A new approach for inclusion and success for students with disabilities*. Washington, DC: A.U. Software, Inc.

Calhoon, M. B., Fuchs, L. S., & Hamlett, C. L. (2000). Effects of computer-based test accommodations on mathematics performance assessments for secondary students with learning disabilities. *Learning Disability Quarterly, 23*(4), 271-281.

Clarkson, P. (1983). Types of errors made by Papua New Guinean students. *Educational Studies in Mathematics, 14*, 355–367.

Clements, M. A. (1980). Analyzing children's errors on written mathematics tasks. *Educational Studies in Mathematics, 11*, 1–21.

Connell, B. R., Jones, M., Mace, R., Mueller, J., Mullick, A., Ostroff, E., et al. (1997). *Principles of Universal Design*. Retrieved January 11, 2005, from http://www.design.ncsu.edu:8120/cud/univ_design/princ_overview.htm

Dawson, L., Venn, M. L., & Gunter, P. L. (2000). The effects of teacher versus computer reading models. *Behavioral Disorders, 25*(2), 105–113.

Dolan, R. P., & Hall, T. E. (2001). Universal Design for Learning: Implications for large-scale assessment. *IDA Perspectives, 27*(4), 22–25.

Dolan, R. P., & Rose, D. H. (2000). Accurate assessment through Universal Design for Learning. *Journal of Special Education Technology, 15*(4).

Ehri, L. C. (1994). Development of the ability to read words: Update. In R. B. Ruddell & M. R. Ruddell (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 323–358). Newark, DE: International Reading Association.

Elliot, S. N., Kratochwill, T. R., & McKevitt, 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.

Elmore, R. F., & Rothman, R. (Eds.). (1999). *Testing, Teaching, and Learning: A Guide for States and School Districts*. Washington, DC.

Farmer, M. E., Klein, R., & Bryson, S. E. (1992). Computer-assisted reading: Effects of whole-word feedback on fluency and comprehension in readers with severe disabilities. *Remedial & Special Education, 13*(2), 50–60.

Fuchs, L. S., Fuchs, D., Eaton, S., Hamlett, C. L., & Karns, K. (2000). Supplementing teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review, 29*(1), 65–85.

Graham, S., & Harris, K. R. (1996). Addressing problems in attention, memory, and executive functioning: An example from self-regulated strategy development. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 349-365). Baltimore, MD: Paul H. Brookes Publishing Co.

Harker, J. K., & Feldt, L. S. (1993). A comparison of achievement test performance of nondisabled students under silent reading and reading plus listening modes of administration. *Applied Measurement in Education, 6*(4), 307–320.

Hebert, B. M., & Murdock, J. Y. (1994). Comparing three computer-aided instruction output modes to teach vocabulary words to students with learning disabilities. *Learning Disabilities Research & Practice, 9*(3), 136–141.

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

Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *Journal of Educational Research, 93*(2), 113–125.

Hollenbeck, K. (2002). Determining when test alterations are valid accommodations or modifications for large-scale assessment. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: validity, technical adequacy, and implementation*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Hollenbeck, K., Rozek-Tedesco, M. A., Tindal, G., & Glasgow, A. (2000). An exploratory study of student-paced versus teacher-paced accommodations for large-scale math tests. *Journal of Special Education Technology, 15*(2), 27-36.

HumRRO. (2003). *CATS Online: Logistic and Construct Evaluation of Computer Administered Assessment* (No. FR-03-40). Alexandria, VA: Human Resources Research Organization.

Johnson, E. S. (2000). The effects of accommodations on performance assessments. *Remedial and Special Education, 21*(5), 261-267.

Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal design and student performance* (No. Technical Report 37). Minneapolis, MN: University of Minnesota, National Center on Education Outcomes.

Landau, S., Russell, M., Gourgey, K., Erin, J., & Cowan, J. (2003). Use of the Talking Tactile Tablet in mathematics testing. *Journal of Visual Impairment and Blindness, 97*(2), 85–96.

Lazer, S. (2002). Personal communication.

Lee, D., & Tindal, G. (2000). *Differential Item Functioning (DIF) as a Function of Test Accommodation*. Delaware Department of Education.

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (Vol. viii, pp. 1–33).

Lundberg, I., & Olofsson, A. (1993). Can computer speech support reading comprehension? *Computers in Human Behavior, 9*, 283-293.

McCullough, C. S. (1995). Using computer technology to monitor student progress and remediate reading problems. *School Psychology Review, 24*(3), 426–439.

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

MetaMetrics. (2004). Lexile Analyzer. Durham, NC: MetaMetrics, Inc.

National Center for Education Statistics. (2002). *NAEP Questions Tool*. Retrieved September 1, 2002, from http://nces.ed.gov/nationsreportcard/itmrls/

Newman, M. A. (1977). An analysis of sixth-grade pupils' errors on written mathematics tasks. In M. A. Clements & J. Foyster (Eds.), *Research in mathematics education in Australia* (Vol. 2, pp. 239–258). Melbourne: Swineburne Press.

Nolet, V., & McLaughlin, M. J. (2000). *Accessing the General Curriculum: Including Students with Disabilities in Standards-Based Reform*. Thousand Oaks, CA:: Corwin Press.

Olson, J. F., & Goldstein, A. A. (1996). *Increasing the Inclusion of Students with Disabilities and Limited English Proficient Students in NAEP* (No. Focus on NAEP v2 n1). Washington DC: National Center for Education Statistics.

Orkwis, R., & McLane, K. (1998). *A curriculum every student can use: Design principles for student access. ERIC/OSEP Topical Brief.* (No. ED423654). Reston, VA: U.S. Department of Education, Office of Special Education Projects.

Parkes, J. T., Suen, H. K., Zimmaro, D. M., & Zappe, S. M. (1999). *Structural knowledge as a pre-requisite to valid performance assessment scores*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Montreal, Quebec, Canada.

Rose, D. H., Meyer, A., Rappolt, G., & Strangman, N. M. (2002). *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria, VA: ASCD Press.

Salyers, F. (2002). Field study of Web-based core content tests goes statewide this fall. *Kentucky Teacher*.

Shepard, L., Taylor, G., & Betebenner, D. (1998). *Inclusion of limited-English-proficient students in Rhode Island's grade 4 mathematics performance assessment* (Technical Report No. 486). Los Angeles: CRESST/UCLA.

Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, Florida: CRC Press.

Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature* (No. Center for Educational Assessment Research Report no. 485). Amherst, MA: School of Education, University of Massachusetts Amherst.

Stanovich, Keith E. Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities, 21*:10, 590–604.

Strangman, N. & Dalton, B. (2005). Using technology to support struggling readers. In D. Edyburn, K. Higgins & R. Boone (Eds.), *Handbook of Special Education Technology Research and Practice.* Whitefish Bay, WI: Knowledge by Design, Inc. (in press).

Swanson, L. (1999). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal of Learning Disabilities, 32*(6), 504–532.

Thompson, S. J., Blount, A., & Thurlow, M. (2002). *A summary of research on the effects of test accommodations: 1999 through 2001* (No. NCEO Technical Report 34). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large-scale assessments* (No. NCEO Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Education Outcomes.

Thompson, S. J., Thurlow, M. L., Quenemoen, R. F., & Lehr, C. A. (2002). *Access to computer-based testing for students with disabilities* (No. NCEO Synthesis Report 45). Minneapolis, MN: University of Minnesota, National Center on Education Outcomes.

Thurlow, M. L., McGrew, K. S., Tindal, G., Thompson, S. J., Ysseldyke, J. E., & Elliott, J. L. (2000). *Assessment accommodations research: Considerations for design and analysis* (No. Technical Report 26). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Tindal, G. & Fuchs, L. (1999). *A summary of research on test changes: An empirical basis for defining accommodations*. Lexington, KY: Mid-South Regional Resource Center/OSEP.

Tindal, G., Glasgow, A., Helwig, R., Hollenbeck, K., & Heath, B. (1998). *Accommodations in large scale tests for students with disabilities: an investigation of reading math tests using video technology*. Washington, DC: Unpublished manuscript with Council of Chief State School Officers.

Tindal, G., Heath, B., Hollenback, K., Almond, P., & Harniss, M. (1998). Accommodating students with disabilities on large-scale tests: An experimental study. *Exceptional Children, 64*(4), 439–450.

Torgesen, J. K. (1993). Variations on theory in learning disabilities. In G. R. Lyon, D. B. Gray, J. F. Kavanagh & N. A. Krasnegor (Eds.), *Better understanding learning disabilities: New views from research and their implications for education and public policies* (pp. 153–170). Baltimore, MD: Paul H. Brookes Publishing Co.

U.S. Department of Education. (2001). *Clarification of the role of the IEP team in selecting individual accommodations, modifications in administration, and alternate assessments for state and district-wide assessments of student achievement*. Washington, DC.

U.S. Department of Education. (2004). *Students with Disabilities to Gain Improved Access to Learning (Press Release)*. Retrieved January 14, 2005, from www.ed.gov/news/pressreleases/2004/07/07272004.html

W3C (2004). Speech Synthesis Markup Language (SSML) Recommendations Version 1.0, from http://www.w3.org/TR/speech-synthesis/

W3C-WAI. (1999). *Web Content Accessibility Guidelines 1.0*, from http://www.w3.org/TR/WCAG10/

Weston, T. J. (1999). *The validity of oral presentation in testing.* Montreal, CANADA: American Educational Research Association.

Weston, T. J. (2003). *NAEP validity studies: The validity of oral accommodation in testing* (No. 2003-06). Washington, DC: U.S. Department of Education – National Center for Education Statistics.

White, S. (2001). *Assessing the Lexile Framework: Results of a Panel Meeting* (No. Working Paper No. 2001-08). Washington, DC: National Center for Education Statistics.

Zuriff, G. E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education, 13*(1), 99–117.

Author Biographies

Robert Dolan is currently Senior Research Scientist at CAST, a not-for-profit education research and development organization that uses technology to make education more flexible and accessible for all students, especially those with disabilities. Dr. Dolan's work focuses on the design, implementation, and evaluation of advanced technology-based learning environments that embed assessments of learner characteristics and progress to continually inform subsequent instruction. Over the past 20 years, he has pursued parallel research and development tracks in cognitive neuroscience and the development of technology-based tools for educational, scientific, and medical research.

As Senior Research Scientist and Instructional Designer at CAST, Tracey Hall specializes in alternative assessment and instructional design grounded in effective teaching practices. Dr. Hall received her Ph.D. in special education from the University of Oregon. Her work has focused on curriculum-based measurement, teacher development, and curricula and instruction for all students.

Manju Banerjee, M.S, M.A., has over 20 years of experience working with and researching students with learning disabilities in the postsecondary setting, and is a certified diagnostician and teacher-consultant for learning disabilities. Ms. Banerjee is currently working on her doctorate at the University of Connecticut.

Euljung Chun is currently working on her doctorate in Special Education at the University of Illinois at Urbana-Champaign. Ms. Chun earned masters degrees in education from Harvard University and the Seoul National University, Korea.

Nicole Strangman, PhD, is Writer, Editor, and Research Publications Coordinator at CAST. A former neuroscientist, Dr. Strangman is interested in topics that intersect neuroscience and education.

J·T·L·A

The Journal of Technology, Learning, and Assessment

www.jtla.org

Technology and Assessment Study Collaborative
Caroline A. & Peter S. Lynch School of Education, Boston College