



Smarter Balanced Assessment Consortium:

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Cara Laitusis, Heather Buzick, Elizabeth Stone, Eric Hansen and Mark Hakkinen
Educational Testing Service
Princeton, NJ

June, 2012





Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Contents

Introduction	4
Audio Presentation for Mathematics	6
Summary of Research on Audio Presentation for Mathematics.....	6
Policy implications.....	7
Implementation considerations.	7
Annotated Bibliography for Audio Presentation for Mathematics	9
Audio Presentation for English Language Arts (ELA)	27
Summary of Research on Audio Presentation for ELA	27
Policy implications.....	28
Implementation considerations.	28
Annotated Bibliography for Audio Presentation for ELA	30
Refreshable Braille for ELA and Mathematics	44
Summary of Research on Refreshable Braille Displays	44
Policy implications.....	44
Implementation considerations.	44
Annotated Bibliography for Refreshable Braille for ELA and Mathematics.....	46
American Sign Language (ASL) for Mathematics.....	49
Summary of Research on ASL for Mathematics	49
Policy implications.....	50
Implementation considerations.	50
Annotated Bibliography for ASL for Mathematics and Reading.....	51
Calculators for Mathematics	60
Summary of Research on Calculators for Mathematics.....	60
Policy implications.....	60
Implementation considerations.	60
Annotated Bibliography for Calculator for Mathematics	61
Writing Tools for English Language Arts	66
Summary of Research on Writing Tools for ELA.....	66
Spell check.	66
Grammar check.....	66

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Word processor.....	66
Speech to text.....	67
Word prediction.....	67
Text to speech (for proofreading.....	67
Policy implications.....	67
Implementation considerations.....	67
Annotated Bibliography for Writing Tools for English Language Arts.....	68
References.....	73

Tables

Table 1: Annotated Bibliography for Audio Presentation of Mathematics.....	21
Table 2: Annotated Bibliography for Audio Presentation for ELA.....	40
Table 3: Annotated Bibliography for Refreshable Braille for ELA and Mathematics.....	48
Table 4: Annotated Bibliography for ASL for ELA and Mathematics.....	57
Table 5: Annotated Bibliography for Calculator Mathematics.....	64
Table 6: Annotated Bibliography for Writing Tools.....	71

Smarter Balanced Assessment Consortium: Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Introduction

Critical to the development of any assessment is the ability to make claims about what students know and can do. The Smarter Balanced Assessment Consortium has made considerable progress in developing content specifications that include accommodation considerations, but policy decisions should be informed not only by content specifications but also by research on the impact of accommodations. The purpose of this literature review is to summarize existing research on controversial and innovative accommodations (and accessibility tools) to inform policy decisions. This document includes an overview of research on the following testing accommodations and accessibility features for students with disabilities:¹

- Audio presentation for mathematics
- Audio presentation for English language arts
- Refreshable braille for mathematics and English language arts
- American sign language for mathematics
- Calculator for mathematics
- Writing tools for English language arts

Some of the studies we reviewed examined validity from the perspective of the interaction hypothesis (see Sireci, Scarpati, & Lee, 2003), which posits that students with disabilities who receive an accommodation that they need will show improvement over a standard administration, whereas scores from students without disabilities will not improve when the students are given the same accommodation. A more relaxed hypothesis is termed *differential boost* (e.g., Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000) and states that an accommodation is appropriate when test scores for students with disabilities are found to improve significantly more (accommodated-standard) than scores for students without disabilities. As an approach for providing validity evidence, both hypotheses can be criticized for not providing any information about predictive validity and for the potential limitation of ceiling effects, which can cause underestimation of the effects of accommodation for high-scoring students.

The document is organized by accommodation/accessibility features and includes a short summary of each feature followed by an annotated bibliography and a summary table. The short summary consists of a description of the feature and a summary of research findings with implications for policy and implementation. For some of the innovative accessibility features (e.g., refreshable braille), the research is very limited and does not provide definitive information on policy issues (e.g., the comparability of scores). In other cases, the research is based on features that were paper based

¹ A second literature review completed by Maria Pennock-Roman, Charlene Rivera, and Lynn Wilner provides an overview of accommodations for English language learners (i.e., glossary, simplified language, audio presentation, translation, and visual representation of linguistic information).



Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

and do not easily inform implementation in a computer-based assessment (e.g., audio presentation by a human proctor). In addition, across all accessibility features, several studies had small sample sizes; we warn readers not to interpret a failure to find a significant result as evidence for no effect of an accommodation, especially when the sample size is small. However, it is our hope that this information will help the Smarter Balanced Assessment Consortia develop sound research-based policies regarding accessibility and testing accommodations.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Audio Presentation for Mathematics

On the basis of our review of the Smarter Balanced Mathematics [Content Specification](#), there does not appear to be an explicit requirement that students read and comprehend math symbols or numbers in visual or tactile form, thereby making audio presentation of numbers and symbols (as well as text) a possible accommodation. This accommodation has the potential to be useful for students with print disabilities (blind, low vision, reading-based learning disabilities) and for English language learners with stronger listening comprehension skills in English than reading comprehension skills. We reviewed 35 studies that evaluated the use of an audio presentation accommodation on mathematics tests with a focus on students with disabilities. Audio presentation can be delivered in a variety of forms such as prerecorded audio, text-to-speech, or an individual teacher reading aloud to a single student or a group of students. This review includes all these approaches to audio presentation; however, very few studies evaluated the impact of text-to-speech or prerecorded audio on a computer-based test. The following information provides guidance for decision making about several forms of audio presentation for mathematics and how this research may inform both policy and practice.

Summary of Research on Audio Presentation for Mathematics

The 25 studies we reviewed took varied approaches, but the majority fell into two categories: (a) studies that focused on analyses of test score gains to detect a differential boost for the appropriate students under the accommodated condition, which would demonstrate that the interaction hypothesis criterion had been met (16 studies), and (b) studies aimed at detecting differential item functioning (DIF) or comparing item difficulty under the accommodated condition relative to the nonaccommodated condition for the focal group of students, which would indicate that performance barriers had been removed (approximately seven studies). Note that the use of DIF to compare groups testing under different accommodation conditions has been supported only when precautions (e.g., experimental design, external matching criterion) have been taken (see, e.g., Buzick & Stone, 2011), and these conditions were not met by many of the included studies employing DIF in this way. Other approaches included prediction equations, evaluation of the internal test structure, and descriptions of test performance using operational data with intact groups, each for assessments taken without and without accommodations.

Many studies did not provide sufficient detail about the groups (e.g., disability groups may have been combined or a specific disability not stated) or the delivery of the accommodation (e.g., what parts of the items were read aloud) to directly interpret accommodation effects. Common technical limitations of the studies we reviewed were small sample sizes for the disability groups and bundling of the audio presentation accommodation with other accommodations. For example, audio presentation is often administered with the allotted testing time extended. The effects of audio presentation in isolation, therefore, are difficult to determine without evaluating progressive accommodations (e.g., no accommodation, extended time only, audio presentation and extended time) in an experimental or quasi-experimental context. Additionally, some studies used off-grade-level content (e.g., using sixth-grade items to test a combined group of sixth and seventh graders), which would not be directly applicable in an accountability context. Still, the studies did provide some evidence to support the use of audio presentation for mathematics for specific student and item types, and they gave some insight into how the accommodation should be implemented to be most effective.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Policy implications.

Of the studies we reviewed, 15 included elementary students, 12 included middle school students, and 4 included high school students, with some studies including more than one grade level. There did appear to be a difference found in the audio presentation effects for the different grade levels. Studies on elementary-level students tended to find beneficial effects, even if small. The studies involving students in the middle and high school grades did not lend as much support for audio presentation. This lack of a benefit for students with disabilities at higher grade levels may be due to several factors. The first involves opportunity to learn, or a Matthew effect: As the student ages, difficulty with reading causes learning deficits in other areas. By middle school, there is a gap in content knowledge that may be indirectly due to a reading deficit. A second factor may be specific to mathematical content. The math-specific vocabulary at the upper grade levels is more rigorous; that is, colloquial mathematical terms at lower grades (e.g., *half*) are supplemented with content-specific terms (e.g., *coefficient*, *hypotenuse*) at higher grades; therefore, the audio presentation accommodation, which assists primarily with decoding, may not help to remove all of the aforementioned barriers at those upper grades. Finally, there is the possibility that reading deficits are not as severe at higher grade levels or that the decoding requirements are less pronounced at higher grade levels.

The studies varied in the groups being evaluated (e.g., students with learning disabilities, students with any type of disability, students with low reading proficiency). Audio presentation of mathematics was shown to be most beneficial to students with low reading fluency who had the mathematical ability to access the item content. The accommodation did not always have a differential boost effect when low readers with or without disabilities were considered. However, for students without the ability to work with the mathematical content, the accommodation had very little benefit. Although various studies suggested that the accommodation could be valid and useful for all students (e.g., removing the decoding requirement could allow all students to focus on problem solving and other math constructs, and the reduced memory load requirements could be beneficial to all), studies that found that there were no or small gains across the groups and that reported negative student reactions to the accommodation are reminders that to be of benefit, the accommodation must be implemented well (e.g., allow choice and self-pacing) and in a mode in which the student has received adequate prior instruction.

Results were not always clear, but benefits of audio presentation were found for the following particular item features: complex items that included numerous verbs, items that required reading and writing, wordy items with nonwordy options, and word problems (vs. computation-only problems).

Because audio presentation of text does not appear to undermine the construct (as articulated in the content specification) and the use of audio presentation does appear to increase scores for some students (i.e., allowing them to show their knowledge, skills, and abilities in mathematics), the use of audio presentation as an accommodation or access tool appears to be warranted. The only caution would be that this feature should be used consistently (until it is no longer needed) to reduce the impact of artificial score changes that would impact year-to-year growth measures.

Implementation considerations.

Because research findings are inconsistent, recommendations for implementation are not clear-cut. In some cases, audio presentation can contribute construct-irrelevant variance to the mathematics construct being measured. For example, the novelty of particular accommodations and their delivery likely had an effect on student performance in many of the studies reviewed. In particular, the use of

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

video storytelling, although seemingly similar to audio, may require instruction in how to extract information rather than entertainment value. However, as the use of gaming and other interactive video and multimedia elements becomes more common in instruction, this concern may be mitigated. Students preferred on-demand accommodations that they could select and control. Studies indicated that recorded audio or text-to-speech is preferable to human readers for various reasons: digital read-aloud could realistically be administered to students individually rather than in a group, the read-aloud pacing could therefore potentially be controlled by the student, and the accommodation would be delivered in a standardized way. However, one use of text-to-speech drew criticism about the level of the voice and the voice quality. Read-aloud delivered by teachers live or via recorded audio controlled by the proctor provoked frustration in some students, particularly students with learning disabilities owing to the pacing. In most studies using recorded audio that was administered in a group setting, all students were given a particular amount of time to answer each item. They never received less than the specified time, and that led to frustration when students finished ahead of time. The use of audio presentation may require additional test changes, such as extended time for all students using audio presentation, to address the extra time requirements that the audio presentation may add.

Regardless of delivery mode, standardized voicing of mathematical content is crucial. This can be directly addressed by following appropriate scripting guidelines. While standardized delivery is always important to ensure that students do not gain information or have information obscured by how the item or option is read, it is even more important for mathematics for several reasons. First, the voicing of symbols must be done using the appropriate math vocabulary so as not to give away information. For example, a simple arithmetic problem like “ $10 - 5$ ” should be read as “ten minus five,” not “ten take away five,” although the latter is the way in which some students are taught to think about subtraction. By vocalizing “take away” rather than “minus,” additional information that may change the construct or the item difficulty is provided. Second, the audio presentation accommodation is not used solely for students with learning disabilities. It has also been offered as an option to students with visual disabilities. For any student, but particularly for students who cannot physically see the problem setup, it is critical to voice the structure of the item appropriately by signposting the key components (such as a fraction with a numerator or denominator). This will allow students to store the item properly in working memory.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Annotated Bibliography for Audio Presentation for Mathematics

- (1) Bielinski, J., Thurlow, M., Ysseldyke, J., Friedenbach, J., & Friedenbach, M. (2001). *Read-aloud accommodations: Effects on multiple-choice reading and math items*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

This study examined DIF for fourth-grade students on multiple-choice mathematics items on Missouri's accountability assessment. Four groups were formed from extant data: a random sample of students without disabilities testing without accommodations ($N = 1,139$), a sample of students without disabilities testing without accommodations with number-correct score distribution matched to the pooled group of students with a primary reading disability with and without read-aloud ($N = 1,006$), students with a primary disability in reading who did not take the test with accommodations ($N = 831$), and students with a primary disability in reading who took the test with read-aloud, possibly bundled with extended time or small-group administration ($N = 1,082$). The latter three groups were each compared with the first group, and read-aloud was delivered via proctors in small-group or individual settings. The analysis approach was a comparison of item difficulty using DIF methodology. Results showed small differences in item difficulty, with slightly less than 20% of the items showing DIF for the group of students with disabilities receiving the read-aloud accommodation relative to nonaccommodated students without disabilities. Of the DIF items, four of six were easier for students who received the read-aloud accommodation. Overall findings and recommendations from the authors were that the read-aloud accommodation did not appear to provide a benefit in this context because there were few items with DIF between students without disabilities and nonaccommodated students with disabilities and there more items showing DIF in students without disabilities versus accommodated students with disabilities. Because of these results, the accommodation did not appear to level the playing field, although the students with disabilities groups in question were naturally formed and may not be considered equivalent. However, the finding that there were DIF items in favor of students with disabilities who received the read-aloud accommodation warranted additional investigation to discover whether that effect was due to the read-aloud removing reading barriers for word problems.

- (2) Bolt, S. E., & Thurlow, M. L. (2007). Item-level effects of the read-aloud accommodation for students with reading disabilities. *Assessment for Effective Intervention*, 33(1), 15–28.

This study examined state test data for Grades 4 and 8 for students with a learning disability who had an Individualized Education Program (IEP) focused on reading. Two groups were compared: students using a read-aloud accommodation (delivered via proctor or audiocassette), possibly with extended time or small-group administration ($N_4 = 1,406$; $N_8 = 1,878$), and students who received no other accommodations other than setting or timing accommodations ($N_4 = 431$; $N_8 = 720$). Comparisons of interest were between the groups on several types of items: math items in general, math items with complex reading required, and easy math items with complex reading required. The analysis design was ANCOVA, in which average scores for each of four item subsets were included as the dependent variables: items that have easy reading level and easy math level, those with easy reading level and hard math level, those with hard reading level and easy math level, and those with hard reading and hard math levels. Reading level was categorized based on numbers of words and syllables, and math level was determined relative to a group of students without disabilities taking the test without accommodations. Performance on computation-only items was included as a covariate in each ANCOVA. In the fourth grade, performance was similar for the group with the read-aloud accommodation and the group without the read-aloud accommodation in both grades. The

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

three-way interaction of accommodation, reading difficulty, and math difficulty was not significant. The interaction of accommodation and item reading difficulty level was significant: students with the read-aloud accommodation performed better on items that were more difficult to read than did students without read-aloud. In eighth grade, results were mixed, with the read-aloud group performing lower than the nonaccommodated group.

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

This DIF study on extant data had the goal of evaluating the hypothesis that more comparable scores were produced in the presence of the read-aloud accommodation (compared with a nonaccommodated administration) on mathematics assessments than on reading/language arts assessments. The math assessment was a multiple-choice state accountability test at the fourth, eighth, and tenth grade levels, and data were available for 3 consecutive years. DIF was undertaken to compare the reference group of randomly sampled, nonaccommodated students without disabilities ($N = 5,000$) to the full populations of (a) nonaccommodated students with disabilities ($N = \sim 6,500$ – $8,000$) and (b) students with disabilities receiving the read-aloud accommodation ($N = \sim 2,000$ – $4,000$). The group with the read-aloud accommodation included students who had also used setting and scheduling accommodations but not response accommodations such as a calculator. Results indicated that moderate to large DIF was present for the group of students taking the mathematics test with read-aloud relative to nonaccommodated students without disabilities. Additional DIF analyses were performed to isolate the read-aloud effect, combining read-aloud with a small-group setting, extended time, or multiple sessions. Overall, 18%–25% of math items in each grade displayed moderate-to-large DIF for the group receiving the read-aloud accommodation. The percentage of DIF items was higher in Grades 8 and 10 relative to Grade 4. The mathematics items flagged for DIF were mainly computation items (in favor of the group with read-aloud) at the upper grades and a mix of computation (in favor of nonaccommodated students without disabilities) and word problems (in favor of the group with read-aloud) at the lower grades. The authors noted that DIF items were also found for the group of nonaccommodated students with disabilities, suggesting that disability was associated with DIF, not just accommodation use.

(4) Burch, M. A. (2002). Effects of computer-based test accommodations on the math problem-solving performance of students with and without disabilities. *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, 63(3), 902.

This quasi-experimental study investigated three accommodation types (computer-read text; video visualization of problem context; and “constructed response,” in which students could input their responses through a mouse or tool bar rather than writing them) for fourth-grade students taking a math performance assessment. The samples included students with reading disabilities ($N = 18$), students with reading and math disabilities ($N = 15$), and a random group of students without disabilities ($N = 16$). In a counterbalanced repeated-measures design, each student took five performance assessments consisting of problem-solving items with authentic contexts. Each performance assessment included a three-paragraph story, a price chart and price pictograph, and four questions to answer based on those materials. Students took the item blocks under standard administration conditions, with each of the three accommodation types separately and with all three accommodations available as a comprehensive administration. Repeated-measures ANOVA was employed to examine differential boost. Results showed evidence of differential boost among the three groups and the five conditions, with mixed results for the two groups of students with

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

disabilities. For students with reading disabilities, results showed increased performance associated with the computer-read text accommodation, but this score boost was smaller than the score boost for students without disabilities. Conversely, students with reading and math disabilities received a larger score boost from the computer-read text accommodation (and from the combined accommodations) than students without disabilities.

(5) Calhoun, 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), 271–282.

This quasi-experimental study involved students with learning disabilities ($N=81$) in 9th to 12th grades taking 3rd-grade-level performance assessments under standard, teacher-read, computer-read, and computer-read with video conditions. The students had average grade level performance of between third and fourth grade on both math and reading, according to teacher ratings. Each student took performance assessments under all four conditions with form and order counterbalanced. Reading and math pretests were administered to categorize students by relative reading proficiency. Analyses were run using ANOVA. Results showed statistically significant score increases associated with each of the three read-aloud conditions relative to the standard administration, with similar scores across the three read-aloud conditions. Students at all reading levels benefited from the read-aloud accommodation. Results also showed that the frequency of rereads was greater with the computer-delivered read-aloud accommodations than with the teacher-read read-aloud accommodation.

(6) Elbaum, B. (2007). Effects of an oral testing accommodation on the mathematics performance of secondary students with and without learning disabilities. *Journal of Special Education*, 40(4), 218–229.

This quasi-experimental study consisted of a repeated-measures administration of two approximately parallel test forms counterbalanced under standard and read-aloud conditions to middle and high school students. It should be noted that the items were not at grade level but rather were targeted to student ability using lower-grade items with difficulty assessed through piloting. The read-aloud condition consisted of a test administrator reading each item twice and allowing students a set amount of time to answer each item. The groups comprised students without disabilities ($N = 134$ for middle school; $N = 103$ for high school) and students with learning disabilities ($N = 187$ for middle school; $N = 201$ for high school). The main analysis was a repeated-measures ANOVA. Findings suggested that both groups benefited from the read-aloud accommodation, but students without disabilities received a higher score boost than students with learning disabilities. However, it was noted that the students without disabilities tended to be low performing in reading, on average. Additional analyses provided evidence that the read-aloud accommodation was more beneficial to students with stronger math skills regardless of whether they had a disability. In addition to their own data analysis, the authors conducted a meta-analysis of empirical research studies on the read-aloud accommodation for math. Results showed that elementary students with learning disabilities experienced a score boost from the read-aloud accommodation on math assessments but that at the secondary level, students without disabilities benefited more from the read-aloud accommodation. On the basis of overall findings, the authors suggested offering the accommodation to all students; however, they cautioned that accommodations offered to all students may introduce construct-irrelevant variance into scores of some students (e.g., owing to pacing).

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

The following two studies report on the same data:

(7a) Elliott, S. N., Kettler, R. J., Beddow, P. A., Kurz, A., Compton, E., McGrath, D., et al. (2010). Effects of using modified items to test students with persistent academic difficulties. *Exceptional Children*, 76(4), 475–495.

(7b) Kettler, R. J., Rodriguez, M. C., Bolt, D. M., Elliot, S. N., Beddow, P. A., & Kurz, A. (2011). Modified multiple-choice items for alternate assessments: Reliability, difficulty, and differential boost. *Applied Measurement in Education*, 24, 210–234.

In these studies, the authors evaluated performance on a modified math assessment that was delivered via computer and consisted of items in the areas of numbers and operations and algebra. Eighth-grade students without disabilities ($N = 256$), students with disabilities who would not be eligible for a modified assessment ($N = 223$), and students with disabilities who would be eligible for a modified assessment ($N = 238$) participated in the study. Each student took three blocks of 13 items each under each of the three conditions—original, modified, and modified with partial reading support (computer-delivered audio read-aloud)—and condition and block order were counterbalanced. One study evaluated performance differences associated with the read-aloud accommodation. Results showed that scores were highest on average for the condition with the read-aloud accommodation for all groups and that students with disabilities did not benefit differentially more than students without disabilities. Within modified assessments, the read-aloud accommodation provided only little benefit in terms of score increase. The other study focused on examining differences in reliability and item difficulty between groups. There were small differences in reliability found across groups. Estimates of item difficulty, averaged across items, were lowest on the modified with read-aloud accommodation administration and highest on the standard administration for students without disabilities and students with disabilities not eligible for the modified assessment. For those students who were eligible for the modified assessment, the estimated average difficulty showed that the modified assessment was less difficult than the modified assessment with the read-aloud accommodation. However, the authors did not perform a statistical test comparing scores between the modified and modified with read-aloud administrations.

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

This study examined teacher judgment of required accommodations versus criteria-based decision making for fourth-grade students without disabilities ($N = 192$) and fourth- and fifth-grade students with learning disabilities ($N = 181$). In the first phase, students took curriculum-based measures in various areas of math under standard, extended time, calculator, and read-aloud conditions. On the basis of whether they received a score boost for each accommodated test, students with learning disabilities were assigned to standard or accommodated conditions on an assessment using items from the Stanford Achievement Test and the Iowa Test of Basic Skills in the second phase. The study found that students with learning disabilities benefited statistically significantly more than students without disabilities when the read-aloud accommodation was provided but only for complex performance assessment tasks that involved reading and writing. Conventional math sections did not show this effect. The findings also appeared to support the idea that students who are more proficient on the construct will benefit more from having construct-irrelevant barriers removed.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

(9) 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.

This quasi-experimental study examined effects of a video presentation of non-computation-only mathematics items at the fifth- and sixth-grade levels for sixth graders of varying reading and mathematics skills. The sample included both students who received mathematics instruction in the general classroom and students who received targeted math assistance and instruction. For the study, students were categorized by their relevant skill levels independently through the use of a basic math skills test and an oral reading fluency prompt. Five groups were formed: low reading–low math ($N = 59$), medium or high reading–low math ($N = 33$), low reading–high math ($N = 35$), medium reading–high math ($N = 35$), and high reading–high math ($N = 35$). The students were each given two approximately parallel math tests consisting of 30 multiple-choice items similar to a state assessment, with counterbalancing. In the video presentation administration, item text was presented on the screen as the words were read by an unseen narrator; students were given a specific amount of time to respond to each item but could move ahead or return at will, although working ahead of the video was not encouraged. Analyses evaluated changes in scores associated with the accommodation (via paired t -tests of standard and video scores), relationships between item attributes and item difficulty, and effectiveness of the video accommodation for six items judged to be complex owing to their syntactical features. The findings suggested that scores were statistically significantly higher with the video accommodation for students with low math skills, regardless of oral reading fluency level. No significant correlations were found between item attributes (e.g., number of words or syllables, complexity of language, math vocabulary requirements) and difference in difficulty between format, with one exception: an increase in the number of verbs per passage appeared to be related to better performance on such items with the video format for low students with low oral reading fluency, particularly if their math skills were high. Overall, results were mixed; very few differences were found between formats, and the low math subgroup benefited from the video accommodation based on analyses with the total test score, whereas the high math–low reading group benefited from the video accommodation when only the complex items were studied. The authors concluded that the video accommodation may not be useful for high oral fluency students; furthermore, it may not be useful on all types of problems.

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

This quasi-experimental study of elementary and middle school students was designed to compare performance on standard and video versions of approximately parallel test forms, with a specific focus on students with reading-based learning disabilities and their performance on complex math problems. Students in Grades 4, 5, 7, and 8 who received special education services and either had an IEP denoting a reading disability or were rated by a teacher as possibly benefiting from a video version of the test were included in the study ($N = \sim 44\text{--}81$ across grades). In each grade, a matched-size random sample of general education students was selected to represent able readers. Students in Grades 4 and 5 took elementary-level forms, and students in Grades 7 and 8 took middle-level forms. All test items were analyzed for difficulty in terms of reading, and four to five items per form were identified as being difficult. Each student was administered two 30-item forms under the two accommodation conditions, and counterbalancing of form and order was carried out. The video version consisted of an unseen narrator reading the test items and options while the words appeared on the video screen. The video was paused for a set length of time after each

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

question to allow for students to respond on their answer sheets. The test booklet in use for that part of the test had only one item on each set of pages to reduce the likelihood of working ahead. The relationship between reading level and test format was evaluated for the difficult items with a repeated-measures ANOVA. Results were mixed. At the elementary level, there was no significant main effect for test format, but there was some evidence that low readers performed better on the accommodated test and able readers on the standard version. At the middle level, low readers did receive a benefit from the read-aloud accommodation over able readers, but there was no significant interaction effect of groups defined by disability status and condition.

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

This quasi-experimental study of elementary and middle school students was designed to evaluate teacher judgment of which students would benefit from a read-aloud accommodation on a math test. Comparisons were also made of performance on standard and video versions of approximately parallel test forms for students with learning disabilities. Students in Grades 4, 5, 7, and 8 who had various disabilities and received special education services ($N = 41\text{--}79$ across grades) and students who took the general education curriculum ($N = 183\text{--}350$ across grades) were included in the study. Students in Grades 4 and 5 took elementary-level forms, and students in Grades 7 and 8 took middle-level forms. Each student was administered two 30-item forms under the two accommodation conditions, and counterbalancing of form and order was carried out. The video version consisted of a split screen with an actor on one side reading the test items and options while the words appeared on the other side. The video was paused for a set length of time after each question to allow for students to respond on their answer sheets, and the time was lengthened if at least one student appeared to need more time complete an item (but time was never shortened). The test booklet in use for that part of the test had only one item on each set of pages to reduce the likelihood of working ahead. Students in special education whose change in scores between the standard version and the video administration met a predefined criterion were the focus of analysis in the study. Teacher judgments of on which administration the students would perform better were correct approximately half the time. Secondary analyses explored the association between pretest reading and math skills and changes in test performance associated with the read-aloud accommodation using a series of *t*-tests. At the elementary level, in fourth grade, students who performed better on the accommodated version also had significantly higher reading and math pretest scores than those who performed better on the standard version. In fifth grade, students who performed better on the standard version had significantly higher math pretest scores than those who performed better on the accommodated version. There were no significant results at the middle level. Additional analyses suggested that students with high math pretest scores performed better on the standard version and that students with low math pretest scores performed better with the read-aloud accommodation. Students with low reading–high math pretest scores, who were hypothesized to be the group to derive the most benefit from the accommodation, performed better on the standard version of the test, on average. One proposed explanation for this outcome was that the accommodation was unfamiliar to some students and proved to be more of a distraction than a help. The overall recommendation was to continue assigning read-aloud as judgment dictated but also to include it in instruction to avoid a novelty effect.

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

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

This study investigated whether the use of a read-aloud accommodation changed the test structure of a high school mathematics exit exam using extant data from tenth-grade students. The factor structures for students without disabilities ($N = 29,137$) and students with mild disabilities ($N = 911$) who took the regular form of the test and students with disabilities requiring an oral accommodation ($N = 934$) taking an alternate form were evaluated. Factor analysis results suggested similar test structure for the three groups. An ANCOVA with demographics and previous reading and math test scores as covariates yielded a significant difference in favor of the students with disabilities who received the read-aloud accommodation versus students with disabilities taking the regular form. The effect size was small, but it represented a substantial improvement in the percentage of students who would pass the exit exam. The main analysis involved using students without disabilities taking the nonaccommodated administration ($N = 89,214$) to develop an equation predicting scores on the exit exam from background variables and previous test scores. This equation was applied to the two groups of students with disabilities (those who received the read-aloud accommodation and those who did not). Students with disabilities who received the read-aloud accommodation scored slightly higher than predicted, on average, and students with disabilities who took the regular administration scored slightly lower than predicted, on average. The authors concluded that the oral accommodation did not change the internal structure of the test and did provide a benefit to the students with disabilities, leveling the playing field.

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

This quasi-experimental study evaluated changes in test performance associated with a read-aloud accommodation on a math performance assessment for fourth-grade students with reading disabilities. Three groups were administered two forms each of a state assessment in a repeated-measures design. The test included both multiple-choice and open-ended questions. Students without disabilities were randomly assigned to either a control group, which was administered both forms without the read-aloud accommodation ($N = 39$), or to a group that took the first form without read-aloud and the second form with read-aloud ($N = 38$). A third group comprised students with reading disabilities who took the first form with read-aloud and the second form without read-aloud ($N = 38$). Note that this design does not account for order effects. Both accommodated groups were compared to the control group using repeated-measures ANOVA. Results suggested that the read-aloud accommodation did not increase scores for students without a disability; however, evidence of a fatigue effect may have confounded these results, or there was not enough power to detect a difference given the small sample size. Students with disabilities performed better with the read-aloud accommodation, but order effects that were not controlled for may have played a role in the score boost. Secondary analyses were conducted by splitting students with disabilities into groups based on reading ability, but there were no significant differences in performance associated with the read-aloud accommodation, and the sample sizes were small. The authors pointed out motivation and test anxiety as additional limitations.

(14) Ketterlin-Geller, L. R., Yovanoff, P., & Tindal, G. (2007). Developing a new paradigm for conducting research on accommodations in mathematics testing. *Exceptional Children, 73*(3), 331–347.

This quasi-experimental study investigated how a read-aloud accommodation delivered via prerecorded audio on a computer-delivered math test relates to item and person characteristics. Third-grade students were classified as lower readers ($N = 33$) and higher readers ($N = 127$) and were administered approximately parallel math test forms under standard and accommodated

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

conditions. Read-aloud ($N_{\text{Lower}} = 17$, $N_{\text{Higher}} = 57$) and simplified language accommodations were randomly assigned to all students. The test was delivered via computer, and for the first form, each student could choose the accommodated or nonaccommodated version of each item. Each item in the first form had an equivalent twin in the second form, so for the second form, the student received each twin in the opposite condition. The items administered in the two test parts were counterbalanced to offset order effects and condition choice effects. Both groups contained some students with disabilities, but the authors made the choice to distinguish accommodation use based on functionality rather than disability status. Test items were categorized based on readability and math difficulty levels, and students performed differently on these item types. Results from several repeated-measures ANOVA analyses were reported. Students with low reading ability scored significantly lower than students with high reading ability, regardless of whether a read-aloud accommodation was used. Separate analysis for lower and higher readers suggested that scores were similar across accommodated and nonaccommodated conditions for subtests defined by different levels of math difficulty and language complexity. Separate analysis with items with high mathematics difficulty and high linguistic complexity showed no significant interaction between accommodation use and reading level, but there was some evidence that lower readers scored higher with the read-aloud accommodation, whereas higher readers scored similarly whether or not they used a read-aloud accommodation. In addition, they found no significant differences in performance related to the read-aloud accommodation for items with low math difficulty or low linguistic complexity. The authors concluded that if students have the math skill required to solve an item, they will not benefit from the use of read-aloud. However, this appears contradictory to some previous studies that showed that students do not benefit from read-aloud unless they have the math skill to access the item once the reading barrier has been removed. The limitations of this study include small sample sizes, no information provided about student choice in using the accommodation for each item and whether it relates to reading ability or math performance, and that groups were defined by a reading pretest instead of disability status.

(15) Lee, D., & Tindal, G. (2000). *Differential item functioning (DIF) as a function of test accommodation*. Newark: University of Delaware Education Research and Development Center.

This study examined DIF for fourth- and fifth-grade students with disabilities using a read-aloud accommodation on state test items at the fourth-grade level. The test was composed of 60 multiple-choice items representing all of the state content strands and was split into two forms that were approximately balanced in terms of difficulty and content. Both forms were administered to each student, once under standard conditions and once with a video-delivered read-aloud accommodation, in a counterbalanced design. In the read-aloud condition, one item at a time was displayed on a video monitor, and the parts of the item were colorized as they were being read. The test booklet in that condition had only one item on each facing set of pages to invoke attention to the item at hand, and students were given a set amount of time (which could be lengthened but not shortened) to respond to each item. All students were allowed to use a calculator on all parts of the test. Students with learning disabilities ($N = 159$) and students without disabilities ($N = 647$) were included in the study. DIF analyses were used to compare administration conditions but not student subgroups. The majority of items classified as wordy (8 of 12) were found to be easier under the video format, relative to the standard administration. Wordy items had mixed results from the different DIF and item difficulty analyses, although a number of them displayed DIF in favor of the

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

video format. However, the authors found that the wordy items that favored the standard condition consistently also had wordy responses.

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

This quasi-experimental study included students from Grades 6, 7, and 8 with reading-based learning disabilities ($N = 62$) and students without disabilities ($N = 198$). The test administered was the Iowa Test of Basic Skills on grade level for each student. Students were randomly assigned to either a standard or a read-aloud condition, and the read-aloud was administered in group settings using a human proctor reading a script. In the read-aloud condition, students could also read along in their test booklets. ANOVA results showed that performance was significantly higher under the administration with the read-aloud accommodation relative to the standard administration, with no significant interaction suggesting that students with disabilities did not benefit from the read-aloud more than students without disabilities. Results are confounded with the influence of extra testing time under the read-aloud administration, which may have inflated scores.

Olson, J. F., & Dirir, M. (2010). *Technical report for studies of the validity of test results for test accommodations: Establishing the validity of test accommodations and score interpretations for students with disabilities: A collaboration of state-based research.*

This report is on validity studies conducted by the Connecticut Department of Education as part of the Connecticut Enhanced Assessment Grant. Two studies, from Connecticut and Nevada, focused on examining differences between students with disabilities and students without disabilities on mathematics tests with and without an audio presentation accommodation with regard to item and test characteristics, test content structure, and differences in test performance.

(17) Connecticut study

A quasi-experimental design was used in which students without disabilities ($N = 366$) and students with disabilities ($N = 282$) took one-half of a seventh-grade math test under standard conditions and the other half with a read-aloud accommodation, with counterbalancing to mitigate any order effects. Read-aloud was delivered via computer with a human voice. The group of students with disabilities comprised students who would be eligible for the read-aloud accommodation, so it included multiple disability subtypes. The reliabilities and item difficulty statistics were, on average, similar for the forms in both conditions. Confirmatory factor analyses provided evidence that the same one factor test structure was plausible for both administration conditions. A repeated-measures ANOVA showed a significant interaction between groups and test conditions, indicating that students without disabilities experienced a slight score decrease, on average, from standard to accommodated, whereas students with disabilities experienced a slight score gain.

(18) Nevada study

Students were administered a seventh-grade math assessment under standard conditions and with a read-aloud accommodation delivered via trained human proctor. Students with disabilities ($N = 212$) experienced small score increases on average on the accommodated administration relative to the standard administration, whereas students without disabilities ($N = 225$) did not receive a score boost from the accommodation, on average. However, results from a repeated-measures ANOVA indicated no evidence of a differential score boost from the read-aloud accommodation. Factor analysis and comparisons of item difficulty showed evidence of similar test structure across standard and accommodated conditions.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

(19) Pomplun, M., & Omar, M. H. (2000). Score comparability of a state mathematics assessment across students with and without reading accommodations. *Journal of Applied Psychology, 85*(1), 21–29.

This study used extant data from a state math assessment for fourth-grade students. Three groups were of interest: a sample of general education students ($N = 1,500$), students with learning disabilities taking the test without an accommodation ($N = 1,369$), and students with learning disabilities who had a read-aloud accommodation ($N = 173$). The goal was to evaluate whether the same construct was being measured for the three groups. Exploratory factor analysis results indicated poor fit of a one-factor model for the two groups of students with learning disabilities. A two-factor model, hypothesized to represent a numerical factor and a writing factor, fit adequately for all three groups. Confirmatory factor analysis results provided some evidence that the same constructs were being measured for students with learning disabilities whether or not they took the test with a read-aloud accommodation, although the covariance between the factors differed across groups. However, the proposed model is one of many that might fit adequately, and the authors concluded that the findings would need to be replicated for other tests to be generalizable. A limitation of the study is that exploratory and confirmatory analyses were carried out on the same data set—confirmatory factor analysis on another sample of students taking the same test would provide stronger evidence of comparability across administration modes.

(20) Schnirman, R. K. (2005). The effect of audiocassette presentation on the performance of students with and without learning disabilities on a group standardized math test. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 66*(6), 2172.

This quasi-experimental study involved sixth- and seventh-grade general education students ($N = 24$) and special education students ($N = 24$) taking parallel test forms of sixth-grade items under standard and audiotaped read-aloud conditions. Counterbalancing was implemented, and a math vocabulary measure was also administered. The participants were all volunteers. The special education students had a specific learning disability and an IEP, did not have emotional or behavioral disorders, and did not require accommodations for a physical disability. ANOVA results indicated that there were no significant differences found for either general education or special education students between the read-aloud and standard conditions for either the math concepts or problem-solving subtests. There was no significant interaction found between group and condition for either subtest. There were no significant differences between students with learning disabilities who performed either high or low on a math vocabulary test, under read-aloud versus standard conditions, on either math concepts or problem solving. The limitations of small sample size and voluntary participation may have contributed to the findings of this study.

(21) Schulte, A. G., Elliott, S. N., & Kratochwill, T. R. (2001). Effects of testing accommodations on standardized mathematics test scores: An experimental analysis of the performances of students with and without disabilities. *School Psychology Review, 30*(4), 527–547.

This quasi-experimental study evaluated changes in math test performance associated with bundled accommodations including the read-aloud accommodation for fourth-grade students without disabilities ($N = 43$) and students with disabilities ($N = 43$). Students without disabilities were randomly assigned a set of accommodations corresponding to a student with disabilities. All students were administered practice materials from the test used in the study to allow students to become familiar with the typical test content and use of accommodations. Each student then took

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

two approximately parallel short forms of the subtest under an accommodated and nonaccommodated condition. Accommodations were bundled in the majority of cases, and the test condition was counterbalanced. ANOVA results showed that though scores for both groups increased under the accommodated condition, the interaction of group and accommodation use was not significant. The failure to find a significant effect could be due to the small sample sizes. The authors did find a significant interaction between group and accommodation status when only the multiple-choice items were analyzed. Additional analyses were performed just for the students receiving the bundle of extended time and read-aloud ($N = 32$ total). For those students, there was no significant interaction of group and condition found, and students with disabilities received a smaller score boost from the accommodations than students without disabilities. Again, small sample sizes may have prevented the authors from finding a significant effect.

(22) Tindal, G. (2002). *Accommodating mathematics testing using a videotaped, read-aloud administration*. Washington, DC: Council of Chief State School Officers.

This quasi-experimental study evaluated the effects of a video-delivered read-aloud to students without disabilities ($N = 575$ in Grades 4–5; $N = 513$ in Grades 7–8) and students with reading-based learning disabilities ($N = 104$ in Grades 4–5; $N = 111$ in Grades 7–8). The read-aloud was delivered in a group format by presenting one item at a time on a video screen at the front of the room, and students were allowed a specified amount of time to respond. Only one item was presented on each set of facing pages to keep students on task with a particular item. The elementary students were administered two parallel 30-item multiple-choice tests assembled from fourth-grade items of a state test. The middle school students were administered two parallel 30-item multiple-choice tests assembled from seventh-grade items of the same state test. Form and condition orders were both counterbalanced, and calculator use was allowed on both forms. ANOVA results differed across grade levels. For the elementary level, the group and condition main effects were both significant, with general education students outperforming students with reading-based learning disabilities and the read-aloud condition producing better scores from both groups than the standard version. There was no significant interaction effect detected. At the middle school level, there were no significant differences in score associated with the read-aloud accommodation for either group. The same results were found at both elementary and middle school levels when a subset of students from both groups who had low reading skills was analyzed separately. The authors found some evidence of form effects, which may have influenced results.

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

This quasi-experimental study investigated a response accommodation (bubbling in an answer sheet vs. responding in the test booklet) and presentation accommodation (students reads test silently to self or teacher reads test in a group setting) for fourth-grade general and special education students. A subset of the analyses focused on performance differences associated with a teacher read-aloud in a group setting relative to the students reading silently to themselves. Students were randomly assigned to either the teacher read-aloud or the student reads silently condition. Focusing on the read-aloud part of the study, several groups who took part in that smaller study were of interest: a random sample of general education students, all special education students participating in the read-aloud study, the 10 general education students in each class ranked as lowest performing by their teachers, the five general education students in each class ranked as lowest performing by their teachers, and the special education students participating in the read-aloud study who also had

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

IEPs in either math or reading. Students were randomly assigned to either the read-aloud or standard format. In the read-aloud condition, the corresponding test booklet page was displayed on an overhead projector for tracking purposes as the proctor read each question twice. ANOVA was used to determine performance differences. Results showed that a random sample of general education students ($N = 66$) scored significantly higher on average with the teacher read-aloud accommodation than special education students ($N = 18$). Focusing on general education students ranked by their teacher to be in the bottom 10 of their class and students in special education with IEPs in math or reading, results suggested that the average score difference between the teacher read-aloud condition and the student reads silently condition was larger, in favor of the teacher read-aloud accommodation, for students with math- and reading-based disabilities compared with low-performing students without disabilities.

(24) Weston, T. J. (2003). *The validity of oral accommodation in testing: NAEP Validity Studies*. Washington, DC: National Center for Education Statistics.

This quasi-experimental study included students with learning disabilities ($N = 65$) and students without disabilities ($N = 54$) in Grade 4 to evaluate changes in performance associated with a read-aloud accommodation. Each student was administered two matched forms of National Assessment of Education Progress (NAEP) mathematics items under standard and read-aloud conditions, and the form and order were counterbalanced. Both word problems and calculation-only problems were included, and the read-aloud accommodation was administered in a group setting with the teacher reading the items aloud. The Grade 3 TerraNova reading test and teacher ratings were also collected. ANOVA results showed a statistically significant increase in scores under the read-aloud condition for all students, on average, and a significantly higher score boost for students with learning disabilities. The significant negative correlation between reading ability and score difference (read-aloud–nonaccommodated) in the combined group of students, and in the group of students with learning disabilities in particular, indicated that less able readers benefited more from the accommodation. Furthermore, all items were easier under the accommodated condition than under the nonaccommodated condition for poor readers with learning disabilities. Score increases associated with the read-aloud accommodation were seen more on word problems than on calculation-only problems for the group of students with learning disabilities. An ANOVA was also performed to investigate whether teachers' assessment of student skills matched better with accommodated or nonaccommodated test performance. The results suggested that teachers' judgments matched the accommodated test better but that teachers' judgments were more predictive for students without disabilities.

Table 1: Annotated Bibliography for Audio Presentation of Mathematics

Study no.	Purpose				Accommodations				Grade			Design			Audio presentation for mathematics		
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Read aloud—text to speech	Read aloud—prerecorded Audio/video	Read-aloud—human	Read-aloud—type not specified	Extra time	Elementary	Middle	High	Experimental/quasi-Experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference
1						✓		✓	✓					✓	1,082	1,139	Small differences in item difficulty between RLD and SWoD, with some items easier for RLD receiving a read-aloud accommodation relative to SWoD not using a read-aloud accommodation
2	✓					✓	✓	✓	✓	✓				✓	1,642 ^a	576 ^a	4th grade: Similar total test performance between RLD with read-aloud and RLD with no read-aloud, with some evidence that the read-aloud accommodation provided a small benefit for items with a higher reading load but were mathematically easy; 8th grade: RLD who received the read-aloud accommodation performed lower than RLD with no accommodations; no other inferences were made because of nonrandom assignment
3		✓							✓	✓	✓		✓		3,000 ^a	6,500 ^a	18%–25% of math items in each grade displayed moderate to large DIF between SWD receiving the read-aloud accommodation and SWoD or SWD who received no accommodations; percentages of DIF items were higher in Grades 8 and 10 relative to Grade 4

Study no.	Purpose				Accommodations				Grade			Design			Audio presentation for mathematics		
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Read aloud—text to speech	Read aloud—pre-recorded Audio/video	Read-aloud—human	Read-aloud—type not specified	Extra time	Elementary	Middle	High	Experimental/quasi-Experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference
4	✓								✓				✓		17 ^a	16 ^a	All students experienced a score boost from the computer-read text; the score boost was largest for RLD/MLD relative to SWoD and RLD only
5	✓				✓	✓						✓	✓		81	N/A	All three types of read-aloud accommodation were similarly associated with score increases relative to the standard administration for RLD; the frequency of rereads was higher with computer-delivered read-aloud vs. teacher-delivered read-aloud
6	✓									✓	✓	✓			194 ^a	119 ^a	Both LD and SWoD benefited from the read-aloud accommodation, but SWoD received a higher score boost than LD; students with higher math skills appeared to benefit more from the read-aloud accommodation
7a,b	✓	✓			✓					✓		✓			238	240	Scores were slightly higher for SWD and SWoD when using the read-aloud accommodation; no evidence of differential boost; reliability was similar across groups, and item difficulty statistics were mixed
8	✓			✓			✓		✓				✓		181	192	SWD benefited more from the read-aloud than SWoD only for complex performance assessment tasks that involved reading and writing; some evidence that students with higher math proficiency benefited more

Study no.	Purpose				Accommodations				Grade			Design			Audio presentation for mathematics		
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Read aloud—text to speech	Read aloud—pre-recorded Audio/video	Read-aloud—human	Read-aloud—type not specified	Extra time	Elementary	Middle	High	Experimental/quasi-Experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference
9	✓	✓								✓			✓		46 ^a	N/A	Among four groups based on math and reading skills, the video accommodation provided a benefit to students with low math skills, regardless of oral reading fluency level, but mixed results suggested that the video accommodation may be useful neither for all types of problems nor for high oral fluency students
10	✓				✓				✓	✓			✓		45 ^a	45 ^a	Elementary level: Some evidence that RLD performed better with the accommodation and SWoD performed better on the standard version, but sample sizes were small; middle level: small benefit for both RLD and SWoD, but the samples sizes were small
11	✓			✓					✓	✓		✓			60 ^a	267 ^a	The accommodation was unfamiliar to some students; teachers were unable to accurately predict which students would benefit from a read-aloud accommodation; SWD and SWoD were included in the study
12	✓	✓				✓					✓		✓		934	29,127	There was evidence of similar test structure among SWD requiring a read-aloud accommodation and students with mild or no disabilities receiving no accommodations; SWD who received the read-aloud accommodation scored higher than predicted

Study no.	Purpose				Accommodations				Grade			Design			Audio presentation for mathematics		
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Read aloud—text to speech	Read aloud—pre-recorded Audio/video	Read-aloud—human	Read-aloud—type not specified	Extra time	Elementary	Middle	High	Experimental/quasi-Experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference
13	✓						✓		✓				✓		38	39	Some evidence of differential boost for RLD relative to SWoD, but samples were small and order effects were not controlled for; secondary analyses on groups based on reading ability provided no evidence of a benefit from the read-aloud accommodation, but sample sizes were small
14	✓	✓							✓				✓		17	57	No evidence of score differences associated with the read-aloud accommodation for low readers vs. high readers, but sample sizes were small
15		✓			✓				✓				✓		159	647	Some evidence that some wordy items were easier with the read-aloud accommodation for both RLD and SWoD
16	✓									✓			✓		62	198	Both RLD and SWoD performed higher with the read-aloud accommodation; no evidence of differential boost
17	✓	✓		✓						✓			✓		282	366	Reliability and item difficulty statistics were similar in read-aloud and standard conditions; there was evidence of similar test structure and evidence of differential boost with SWDs receiving a score gain and SWoD receiving a score loss associated with the read-aloud accommodation

Study no.	Purpose				Accommodations				Grade			Design			Audio presentation for mathematics		
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Read aloud—text to speech	Read aloud—pre-recorded Audio/video	Read-aloud—human	Read-aloud—type not specified	Extra time	Elementary	Middle	High	Experimental/quasi-Experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference
18	✓		✓			✓				✓			✓		212	225	Evidence of similar test structure; small score differences; no evidence of differential boost for SWD and SWoD
19			✓				✓		✓					✓	173	1,435	Some evidence that the test structure was the same for RLD with a read-aloud accommodation, SWoD, and RLD with no accommodations, but findings need to be replicated because exploratory and confirmatory analyses were conducted on the same samples
20	✓				✓					✓			✓		24	24	No evidence of differences in performance associated with the read-aloud accommodation for RLD and SWoD, but sample sizes were small
21	✓								✓				✓		43	43	The bundled accommodations were associated with higher scores for both SWD and SWoD; in a secondary analysis on a small sample of students receiving only read-aloud and extra time, SWD received a smaller score boost than SWoD
22	✓								✓	✓			✓		108 ^a	544 ^a	Grades 4-5: Score boost for RLD and SWoD, no differential boost; Grade 7-8: No evidence of score boost
23	✓					✓			✓				✓		18	66	Evidence of differentially higher scores for SWD with read-aloud relative to low-performing SWoD

											Audio presentation for mathematics							
Study no.	Purpose				Accommodations				Grade			Design			Sample size focal	Sample size reference	Major findings	
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Read aloud—text to speech	Read aloud—pre-recorded Audio/video	Read-aloud—human	Read-aloud—type not specified	Extra time	Elementary	Middle	High	Experimental/quasi-Experimental	Operational data				Qualitative/descriptive
24	✓						✓		✓				✓			65	54	Read-aloud was associated with increased scores for both RLD and SWoD, with a differentially higher increase, on average, for RLD

Note. RLD = students with reading-based learning disabilities; MLD = students with math-based learning disabilities; LD = students with learning disabilities; SWD = students with any type of disability or unspecified subtypes; SWoD = students without disabilities.

^aStudy has more than one focal and/or reference group, within or across grades. Sample sizes are averaged.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Audio Presentation for English Language Arts (ELA)

Based on the draft *Content Specifications for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects*, published by Smarter Balanced in September 2011, it appears that the Smarter Balanced assessment will define reading to include audio presentation after a certain grade. This assumption is based on the Explication of Claim 1 (students can read closely and critically to comprehend a range of increasingly complex literary and informational texts), which states that

providing assistive technologies such as text to speech may not be considered appropriate up through an intermediate-level grade, say, four or five. After that, the use of text to speech (or a human reader) is considered an appropriate avenue of access to allow students to demonstrate that they are able to read closely and critically to comprehend a range of increasingly complex literary and informational texts. (Hess, 2011, p. 27)

Though this specificity in the draft content specifications is important, there is likely to be disagreement among Smarter Balanced states on how text-to-speech accommodations should be implemented in higher grades. Similarly, there may be disagreement on other issues related to text-to-speech accommodations. The following information provides guidance for decision making about several forms of audio presentation (including text-to-speech) on the Smarter Balanced English Language Arts and Literacy summative assessment.

Summary of Research on Audio Presentation for ELA

We reviewed 21 studies that evaluated audio presentation for students with disabilities on ELA assessments. In all cases, the primary focus of the study was on either a broad construct of “English language arts” or the specific construct of reading comprehension. Nearly all studies reviewed focused on students with learning disabilities or groups of students with disabilities (in general, consisting of large numbers of students with learning disabilities). However, we included one recent publication that provides a compelling argument for audio-supported reading (text-to-speech along with braille or magnification) as a crucial literacy skill for students with visual impairments. None of the studies specifically evaluated the impact of audio presentation on other aspects of ELA such as writing prompts, grammar, listening, or speaking. One study on text-to-speech for proofreading, however, is summarized in the following section on writing tools.

Of the 21 studies we reviewed, several were based on the same data set, so results are limited to 17 unique data sets. In most cases, the additional studies on the same data sets examined the impact of factor structure or DIF as well as differential performance gains between groups of students. Evidence of similar factor structure was found in five of the six studies that used confirmatory factor analysis. Three of the four studies that examined differential functioning analyses or compared item difficulty showed evidence of differences in item difficulty, particularly in elementary grades.² These findings may have implications for selection of items on a computer-adaptive assessment.

A total of 11 studies evaluated differential boost. Six studies included students in the elementary grades; all these studies found evidence of a differentially higher boost for students with disabilities

² The findings of differential item functioning may be due to study design. As described by Buzick and Stone (2011), using groups that differ in their administration condition (accommodated vs. nonaccommodated) violates the assumptions of differential item functioning analyses and may lead to the finding of more items exhibiting differential item functioning.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

relative to students without disabilities, with small to modest score increases found for nondisabled students. All but one of these studies evaluated a read-aloud accommodation for the entire test, including passages. Seven differential boost studies included middle school students; two studies that used a repeated-measures design found support for the interaction hypothesis, whereas two studies with random assignment to a single testing condition found high scores on average for students receiving the read-aloud accommodation, regardless of disability status, with no evidence of differential boost. One of the latter studies included a large amount of extra time only for the read-aloud administrations, which could have inflated the score increases for all students. Two of the three differential boost studies that found no evidence of score differences associated with audio presentation for middle school students had very small sample sizes, which may have prevented the authors from finding an effect.

Across the studies, there were differences in how the audio presentation was administered, what content was read aloud (in at least eight studies, the entire test was read aloud, including passages), and the types of students included in the study (e.g., students with reading-based learning disabilities or a combined group of students with various types of disabilities). In terms of how the audio presentation was administered, the studies included human readers, prerecorded audio (via computer, CD, and MP3 players), prerecorded text-to-speech, and text-to-speech. Studies that evaluated a computer-delivered audio presentation accommodation (text-to-speech or prerecorded text-to-speech) found that the quality of the speech and students' prior experience using text-to-speech were major factors in the impact of this feature.

Policy implications.

The most important consideration for setting policy on audio presentation accommodations for the Smarter Balanced English Language Arts assessment is the definition of the construct being measured on the test at different grade levels and/or items and tasks designed to measure specific claims. Additional considerations include defining the parts of the test that are permitted to be read aloud (i.e., passages, questions, some items, all items, student chooses the items) and who is permitted to use the accommodation (e.g., all students, students with disabilities defined by IEP/504 plans, based on a reading subtest, chosen by the teacher). The results from the studies reviewed tend to support the use of a read-aloud accommodation when decoding is not a part of the construct being measured and in middle school if the read-aloud accommodation is offered without significantly extending the testing time.

Implementation considerations.

There are several areas to consider when deciding how to implement an audio presentation accommodation that can impact its use and efficacy and, ultimately, the validity of inferences based on scores. One decision is the use of text-to-speech, prerecorded audio, recorded text-to-speech, or human readers. One consensus from the technical advisors was that any form of audio presentation for the English language arts assessment should include text-to-speech because it maximizes the independence of students and is a critical career and college readiness skill for students with print disabilities. There was no consensus, however, on the exact grade level for inclusion of text-to-speech as a test feature. Another advantage of text-to-speech (and some forms of prerecorded text-to-speech) is that it allows the test taker to control the pace of the audio and integration of audio with other accommodations such as synchronized highlighting. When the student does not control the pace of the test administration with a read-aloud accommodation, the result can be too much testing time, which can lead to either fatigue or extra time as an additional, unintended accommodation that can alter test scores. Among the studies reviewed here, teacher-paced read-

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

aloud tended to add a nuisance dimension to the test administration, particularly for students without disabilities in the studies.

Cuing is a potential source of construct-irrelevant variance that can occur when the read-aloud accommodation is delivered via human proctor. Computer-delivered audio presentation or audio delivered via MP3 player or CD provided the student with the option to request a reread with benefits over audio presentation via human proctor in a group setting. However, the use of prerecorded audio (CD, MP3, and some computer-delivered audio) is that it does not allow students to navigate at the word (or phrase) level owing to the track features on both CDs and most MP3 players. However, the studies on synthetic speech reviewed here showed that audio presentation delivered via text-to-speech may introduce a nuisance dimension (e.g., construct-irrelevant variance) relative to a prerecorded human voice. Training in the computer-based testing environment and with a computerized read-aloud accommodation is recommended.

In terms of deciding which students receive the audio presentation accommodation (e.g., which subset of students with learning disabilities), teacher judgments have been shown to be inaccurate, but a reading-pretest or multistage assessment to route students to the accommodated assessment has shown promise. Because there are several implementation considerations, any decision on how to implement audio presentation accommodations should be coupled with continued monitoring and validation research. For example, evaluating the relationship between test scores and measures of college and career readiness for students receiving the accommodation and those who do not can provide evidence about whether meaning of test scores is the same under accommodated and nonaccommodated test administrations.

Annotated Bibliography for Audio Presentation for ELA

- (25) Bielinski, J., Thurlow, M., Ysseldyke, J., Freidebach, J., & Freidebach, M. (2001). *Read-aloud accommodation: Effects on multiple-choice reading and math items* (Technical Report No. 31). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

This study used extant data from a third-grade reading assessment from the Missouri Assessment Program to examine whether the reading aloud of test passages and items by a human proctor altered the construct being measured by the test. There was a control group of nonaccommodated students without disabilities ($N = 1,002$) and three focal groups each taking the test once—the main focal group comprised students with a reading disability who received a read-aloud accommodation either alone or in combination with extended time, small-group administration, or both ($N = 661$), whereas the other two groups comprised students without disabilities who were low performing ($N = 995$) and students with a reading disability who took the assessment without any accommodations ($N = 600$). Average test performance across the three focal groups was similar. The analysis approach was a comparison of item difficulty using DIF methodology. The results suggested that the accommodation(s) introduced construct-irrelevant variance to the test scores (increased item difficulty and more items exhibiting DIF). This study provided evidence that the third-grade reading assessment scores did not mean the same thing when the test is taken with and without a read-aloud accommodation. Limitations of the study include nonrandom assignment, the possibility that the accommodation was not administered uniformly or properly to students, and that DIF statistical methodology that assumed equivalent administration conditions across groups was used to compare groups that differed in administration conditions (see Buzick & Stone, 2011).

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

This study used extant data from an unidentified state reading assessment for students in Grades 3, 7, and 11 to evaluate whether the same construct was being measured for students who received a read-aloud accommodation via human proctor as for those who did not receive an accommodation. In each grade, there was a control group of nonaccommodated students without disabilities ($N = 5,000$), a focal group of nonaccommodated students with disabilities ($N = 5,000+$), and a focal group of students with disabilities who received read-aloud accommodation with or without additional setting and scheduling accommodations ($N = 2,000+$). The analysis approach was DIF methodology. Results showed a large percentage of DIF items for students receiving the read-aloud accommodation, particularly relative to the percentage of DIF items found on the math assessment for students taking the assessment with a read-aloud accommodation and also relative to student with disabilities who did not receive accommodations on the reading assessment. The percentage of DIF items was found to decrease slightly for higher grade levels. DIF items that were relatively easier for students with the read-aloud accommodation were mostly reading items; items that were relatively more difficult for students with the read-aloud accommodation were mostly writing items.

- (27) Cook, L., Eignor, D., Sawaki, Y., Steinberg, J., & Cline, F. (2010). Using factor analysis to investigate accommodations used by students with disabilities on an English-language arts assessment. *Applied Measurement in Education, 23*(2), 187–208.

This study used extant data from a state ELA assessment taken by students in Grade 4 to evaluate the similarity of test structure between the standard administration and accommodated administrations. There were four groups: nonaccommodated students without disabilities and three Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

groups of students with learning disabilities (no accommodations, accommodations based on 504 plans/IEPs, and read-aloud accommodation). The sample size for all groups was 500. Results from confirmatory factor analysis support a similar factor structure for all groups.

(28) Crawford, L., & Tindal, G. (2004). Effects of a read aloud modification on a standardized reading test. *Exceptionality, 12*, 89–106.

This study used five reading passages that were developed as part of a large state assessment (in North Carolina or Oregon), administered to students in Grades 4 and 5 from a nonrandom sample of schools in North Carolina and Oregon, to evaluate the impact of a read-aloud accommodation on reading test scores. A repeated-measures design was used with all students taking both a standard version of the assessment and a parallel form administered with a read-aloud accommodation delivered via video of a human proctor. There were no performance differences across grades ($N = 74$ for fourth graders, 264 for fifth graders), so results were combined in the analyses. The sample included three groups of students from two states: students with disabilities ($N = 76$, the majority with learning disabilities), students without disabilities ($N = 173$), and students receiving Title I services ($N = 89$). Repeated-measures ANOVA was the method of analysis. Results showed that performance was significantly higher with the read-aloud accommodation for all students and differentially higher for students with disabilities (effect size = .71 for students with disabilities; .28 for students without disabilities), including a higher percentage of students with disabilities receiving a significant score boost from the read-aloud accommodation relative to students without disabilities. Additional results suggested that teachers overidentified the number of students who would receive a significant score boost and underestimated performance on the standard administration (i.e., no read-aloud). The authors recommended evaluating the need for read-aloud on an individual basis, based on the fact that the score boost for all students suggested that the construct being measured was altered by the presence of the read-aloud accommodation. The authors also found differences in performance across the two parallel test forms that were administered with the read-aloud accommodation but no difference across forms under the standard administration, suggesting that the read-aloud accommodation may have interacted with individual test items in different ways.

The following two studies report on the same data:

(29a) Elliott, S. N., Kettler, R. J., Beddow, P. A., Kurz, A., Compton, E., McGrath, D., et al. (2010). Effects of using modified items to test students with persistent academic difficulties. *Exceptional Children, 76*(4), 475–495.

(29b) Kettler, R. J., Rodriguez, M. C., Bolt, D. M., Elliott, S. N., Beddow, P. A., & Kurz, A. (2011). Modified multiple-choice items for alternate assessments: Reliability, difficulty, and differential boost. *Applied Measurement in Education, 24*, 210–234.

These studies focused on a modified reading comprehension and vocabulary assessment developed by Discovery Education Assessment, administered to eighth-grade students from four states (Arizona, Hawaii, Idaho, and Indiana). students without disabilities ($N = 256$), students with disabilities not eligible to take a modified assessment ($N = 228$), and students with disabilities who would be eligible for a modified assessment ($N = 237$) took the assessment under three conditions—original, modified, and modified with read-aloud—with each item taken under only one condition. The read-aloud accommodation was delivered via recorded voice on a computer-based test, with each item appearing on a single screen. A recorded voice read item directions and stems automatically.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Item options and graphics that contained words could also be played aloud by clicking on an audio file icon. Read-aloud was not permitted on some parts of reading items, for example, key vocabulary words were not able to be read aloud. Students were randomly assigned to different item-condition orders. The sample comprised students with different disability subtypes, with the largest subgroup containing students with learning disabilities and a higher percentage of students with mental retardation in the group eligible for the modified assessment. Coefficient alpha was similar across all three conditions. One study compared Rasch difficulty parameter estimates, and the other compared total scores. Rasch difficulty parameter estimates, averaged across items, were lowest on the modified with read-aloud accommodation administration and highest on the standard administration for students without disabilities and students with disabilities not eligible for the modified assessment. For those students who were eligible for the modified assessment, the estimated average difficulty showed that the modified assessment was less difficult than the modified assessment with the read-aloud accommodation. However, the authors did not perform a statistical test comparing scores between the modified and modified with read-aloud administrations. Scores were highest for all groups in the modified with read-aloud condition, but effect sizes were negligible for the modified assessment with read-aloud versus the modified assessment with no read-aloud for all groups. In addition, no significant interaction was found between groups and testing conditions, indicating that students with disabilities did not receive a differential boost from the read-aloud accommodation. However, the incremental effect size was slightly larger for students with disabilities (.11) relative to students without disabilities (.01).

(30) Fletcher, J. M., Francis, D. J., O'Malley, K., Copeland, K., Mehta, P., Caldwell, C. J., et al. (2009). Effects of a bundled accommodations package on high-stakes testing for middle school students with reading disabilities. *Exceptional Children*, 75(4), 447–463.

This study used a practice form of the Texas Assessment of Knowledge and Skills reading test, administered to students in Grade 7 from 17 middle schools in suburban southeast Texas, to evaluate changes in test scores associated with a read-aloud accommodation on an assessment administered in two testing blocks either on 1 day or on 2 successive days. Students with word reading disabilities from schools that serve students with dyslexia or who received special education services were matched with average readers based on gender and ethnicity in the sampling procedure. Read-aloud was delivered via human proctor, with questions and answer choices read aloud along with proper nouns. Students were randomly assigned to one of three testing conditions: standard administration, read-aloud with 1-day administration, or read-aloud with 2-day administration. All conditions had a small group setting, read-aloud of questions and answer choices and proper nouns, read-aloud by human proctor, read-aloud with 1-day administration, read-aloud with 2-day administration, standard administration, matched students with word reading disabilities with an average reader based on gender and ethnicity, and randomly assigned pairs to one of three testing conditions. One group comprised mostly students with reading disabilities in special education, with the remaining students having been diagnosed with dyslexia ($N = 168$), and the other group comprised average readers ($N = 191$). Results suggested no evidence of differential boost, with an interaction that was not statistically significant. Across all groups, the performance of students with the 2-day administration was highest on average, and average performance with the standard administration was lowest. Covariates that included various external measures of reading ability were found not to be related to performance differences associated with the accommodations. There was some evidence that average readers performed poorly on the experimental version of the test, which the authors argue may have influenced the finding of no significant interaction.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

- (31) Fletcher, J. M., Francis, D. J., Boudousquie, A., Copeland, K., Young, V., Kalinowski, S., et al. (2006). Effects of accommodations on high-stakes testing for students with reading disabilities. *Exceptional Children*, 72, 136–150.

This study evaluated the role of reading aloud by a human proctor of proper nouns, comprehension stems, and possible responses in performance on a practice form of the Texas Assessment of Knowledge and Skills reading test for students in Grade 3 with dyslexia and poor decoding skills. Students with dyslexia ($N = 91$) and a control group of students without disabilities with average reading ability ($N = 91$) were randomly assigned to an accommodated administration or a standard administration. The accommodated administration also included extra time via two block testing sessions and small group setting. Students in both conditions read the passages independently. ANCOVA was used with a multilevel model to account for nesting of students within classrooms and districts. Results showed a large increase in performance associated with the read-aloud accommodation for students with dyslexia (effect size = .91) and poor decoding skills relative to those without (effect size = .15), with no significant change in performance for students in the control group, on average, across standard and accommodated administrations. There was also a significant increase in the odds of passing for accommodated students with dyslexia and poor decoding skills, with no effect found for the control group.

- (32) Flowers, C., Kim, D.-H., Lewis, P., & Davis, V. C. (2011). A comparison of computer-based testing and pencil-and-paper testing for students with a read-aloud accommodation. *Journal of Special Education Technology*, 26, 1–12.

This study used extant data from a southeastern state to compare reading test scores from a large-scale state assessment across two administration modes: a computer-based test (CBT) with read-aloud delivered via digital text readable with text or screen reader and a paper-and-pencil test (PPT) with read-aloud delivered via human proctor in a one-to-one setting. Additional accommodations in the CBT setting included different screen and text presentation modes, one question per screen, choice of formats for reading passages, alternative text for graphics, and headphones. Students with disabilities who were eligible for the read-aloud accommodation in Grades 3–11 took the PPT ($N = 2,940$ – $3,409$ across grades) or the CBT ($N = 33$ – 303 across grades). Comparisons of descriptive statistics across the nonrandom groups showed higher scores on average for the PPT with read-aloud via human proctor condition relative to the CBT with read-aloud. Standardized mean differences were similar across grades. The authors also matched seventh- and eighth-grade students in the CBT group to students in the PPT group via propensity score matching based on reading and math test scores and free or reduced lunch status. For these matched students, scores for the PPT administration were significantly higher than for CBT, with a larger effect size in Grade 7 (.65) relative to Grade 8 (.34). Results from DIF analyses for the full sample were inconclusive. Student surveys suggested that students favored the CBT administration.

- (33) Harris, L. W. (2008). *Comparison of student performance between teacher read and CD-ROM delivered modes of test administration of English language arts tests* (Doctoral dissertation). Retrieved from Dissertations & Theses: Full Text database. (Publication No. AAT 3321402)

This study used extant data from the Palmetto Achievement Challenge Test reading assessment for students in South Carolina in Grades 6, 7, and 8. Factor structure and performance were compared across two different read-aloud administrations: read-aloud of an oral script by a human proctor and computer delivery of a human proctor reading an oral script via CD-ROM. The CD-ROM allowed students to select directions, questions, or answer choices to be read aloud. The sample comprised

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

groups of students with learning disabilities only and groups of students with any other primary disability or multiple disabilities including a learning disability. More than 1,500 students per grade took the assessment with a read-aloud accommodation delivered via proctor, and approximately 300 students per grade took the assessment with the read-aloud accommodation delivered via CD-ROM. All groups comprised more students with learning disabilities than those with multiple disabilities (approximately twice as large). Confirmatory factor analysis and MANCOVA were used to compare the two modes of administration. The author found no significant differences in performance across modes and evidence of measurement equality across modes in two of the three grades studied.

(34) Huynh, H., & Barton, K. (2006). Performance of students with disabilities under regular and oral administrations of a high-stakes reading examination. *Applied Measurement in Education*, 19(1), 21–39.

This study compared test factor structure and performance for 10th-grade students with disabilities taking the South Carolina high school exit exam in reading with or without a read-aloud accommodation. Extant data were used with test scores from three groups: students with disabilities receiving a read-aloud accommodation ($N = 822$), students with disabilities who did not receive a read-aloud accommodation ($N = 3,022$), and students without disabilities who did not receive a read-aloud accommodation ($N = 85,457$). The two groups of students with disabilities comprised multiple disability subtypes, primarily learning disabilities. The read-aloud accommodation was either prerecorded audio operated by administrator or student or read aloud by a human proctor. The test form administered with the read-aloud accommodation was parallel to the regular forms, except that items with onomatopoeias were replaced with items with similar content and difficulty. Students with disabilities may have received other accommodations, and the test was not timed. Confirmatory factor analysis suggested that the factor structure of the test was similar across all three groups. ANCOVA results suggested that test scores, after controlling for eighth-grade reading performance and demographic variables, were similar for students with disabilities taking the test with the read-aloud accommodation and students with disabilities who did not receive a read-aloud accommodation.

(35) Kosciolk, S., & Ysseldyke, J. E. (2000). *Effects of a reading accommodation on the validity of a reading test*. Retrieved from <http://education.umn.edu/NCEO/OnlinePubs/Technical28.htm>

This study evaluated changes in test performance associated with a read-aloud accommodation delivered via prerecorded audio for students in Grades 3–5 in a suburban school district taking the California Achievement Tests reading comprehension test. Under a counterbalanced, repeated-measures design, students with disabilities ($N = 15$) and students without disabilities ($N = 17$) were administered two parallel forms of a reading comprehension test delivered with and without the read-aloud accommodation, with both administrations in a small-group setting. Findings included no significant differences in test scores between accommodated and nonaccommodated conditions for either group. Repeated-measures ANOVA revealed no interaction effect. Questions about student preferences for the read-aloud accommodation revealed a higher percentage of students with disabilities favoring the read-aloud accommodation relative to students without disabilities.

The following three studies are from the Designing Accessible Reading Assessments (DARA) grant:

(36a) Laitusis, C. C. (2010). Examining the impact of audio presentation on tests of reading comprehension. *Applied Measurement in Education*, 23, 153–167.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

This study explored changes in test performance associated with a read-aloud accommodation using a sample of fourth- and eighth-grade students from 84 public and private schools in New Jersey. Students with reading-based learning disabilities ($N = 527$ in fourth grade; $N = 376$ in eighth grade) and students without disabilities ($N = 654$ in fourth grade; $N = 471$ in eighth grade) took the reading comprehension subtest of the Gates–McGinitie Reading Tests, Fourth Edition, with and without a read-aloud accommodation administered via prerecorded audio delivered on CD with headphones. Extra time and answers recorded in the test booklet were additional accommodations that were offered under both conditions. The standard and read-aloud administrations were counterbalanced across schools, with two parallel test forms administered in random order to students within schools. Repeated-measures ANOVA showed a significant interaction between disability and the use of the read-aloud accommodation in Grade 4 (effect size = .57 for students with reading-based learning disabilities; effect size = .14 for students without disabilities) and in Grade 8 (effect size = .32 for students with reading-based learning disabilities; effect size = .06 for students without disabilities). When controlling for reading fluency and ceiling effects, the authors still found evidence of differential boost.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

- (36b) Cook, L., Eignor, D., Steinberg, J., Sawaki, Y., & Cline, F. (2009). Using factor analysis to investigate the impact of accommodations on the scores of students with disabilities on a reading comprehension assessment. *Journal of Applied Testing Technology*, 10(2). Retrieved from <http://www.testpublishers.org/assets/documents/Special%20issue%20article%203.pdf>

This study used the fourth-grade sample of students with and without reading-based learning disabilities described in Laitusis (2010). Factor analysis using item parcels suggests the same factor structure whether the Gates–McGinitie reading test was given with or without a read-aloud accommodation for both students with and without reading-based learning disabilities. While not reported in the study, the authors pointed out that they found similar results for the sample of students in eighth grade.

- (36c) Middleton, K. V. (2007). *The effect of a read-aloud accommodation on items on a reading comprehension test for students with reading-based learning disabilities* (Unpublished doctoral dissertation). University of Iowa, Iowa.

This study used DIF methodology to compare item characteristics across standard and read-aloud administrations for the subset of students with reading-based learning disabilities. Results suggested that items administered under the accommodated condition were relatively easier. In a second analysis, students without disabilities who did not receive an accommodation were compared to the two groups of students with reading-based learning disabilities. Results suggested that a different construct was measured when the test was administered with a read-aloud accommodation relative to the standard administration.

- (37) McKeivitt, B., & Elliott, S. N. (2003). Effects and perceived consequences of using read-aloud and teacher-recommended testing accommodations on a reading achievement test. *School Psychology Review*, 32, 583–600.

This study evaluated changes in test scores associated with a read-aloud accommodation bundled with teacher-recommended accommodations for eighth-grade students from one school in a midwestern suburban city taking the TerraNova Multiple Assessments Reading test. Students with reading-based learning disabilities ($N = 40$) and students without disabilities ($N = 39$) were randomly assigned to one of two conditions: (a) no accommodations–teacher-recommended accommodations or (b) no accommodations–read-aloud accommodation delivered via prerecorded audio plus teacher-recommended accommodations. Students in both conditions received the standard administration first and the accommodated administration in a small-group setting second. Repeated-measures ANOVA separately for each condition resulted in no significant interactions between accommodation use and presence of a disability. The authors found no significant score gain for the read-aloud accommodation, although there were significant score gains on average for read-aloud plus teacher-recommended accommodations relative to no accommodations for all students (but the gain was within the range of true scores). A higher percentage of students with disabilities had positive opinions about the read-aloud accommodation than students without disabilities, but overall, feelings were mixed. Study limitations include offering the accommodated administration second to all students and combining students with disabilities who may or may not have needed a read-aloud accommodation.

- (38) 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, 248–255.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

This study evaluated changes in test performance associated with a read-aloud accommodation using students in Grades 6–8 in two midwestern schools. Students with a reading-based learning disability ($N = 62$) and students without disabilities ($N = 198$) took the Iowa Tests of Basic Skills reading comprehension subtest and were randomly assigned to either a standard administration or an accommodated administration with read-aloud delivered by human proctor. Results suggest that performance was significantly higher under the administration with the read-aloud accommodation relative to the standard administration, with no significant interaction suggesting that students with disabilities did not benefit from the read-aloud accommodation more than students without disabilities. Results are confounded with the influence of extra testing time under the read-aloud administration, which may have inflated scores.

Olson, J. F., & Dirir, M. (2010). *Technical report for studies of the validity of test results for test accommodations—Establishing the validity of test accommodations and score interpretations for students with disabilities: A collaboration of state-based research.*

This report is on validity studies conducted by the Connecticut Department of Education as part of the Connecticut Enhanced Assessment Grant. Two studies, from Connecticut and Kentucky, focused on examining differences between students with disabilities and students without disabilities on reading comprehension tests with and without an audio presentation accommodation with regard to item and test characteristics, test content structure, and differences in test performance.

(39) Connecticut

This study explored changes in test structure and test performance associated with a read-aloud accommodation administered via a digital voice text reader for seventh-grade students in 18 public schools in Connecticut taking a computer-based reading comprehension test similar to the state ELA assessment. Students with disabilities who were deemed eligible for a text reader accommodation ($N = 206$) and students without disabilities, matched by demographic characteristics ($N = 200$), took the test both under a standard administration and under the accommodated administration. Item statistics and confirmatory factor analysis results showed a similar test structure across the two administrations. Results from repeated-measures ANOVA suggested that scores did not change significantly across administrations for either students with disabilities or for those without. The authors pointed out a major limitation in the study: poor quality of the digital text reader.

(40) Kentucky

This study explored changes in test structure and test performance associated with a read-aloud accommodation on a reading test similar to the Kentucky state assessment administered via human proctor. Fourth-grade students with disabilities who were eligible for the read-aloud accommodation ($N = 150$) and students without disabilities matched with background characteristics ($N = 145$) took the test under a standard administration and with the read-aloud accommodation. Results suggested that items were easier and less reliable when the test was administered with the read-aloud accommodation, but evidence suggested that one factor was being measured by both administration conditions. Repeated-measures ANOVA showed that scores were significantly higher for all students and that students with disabilities benefited more than students without disabilities.

(41) Randall, J., & Engelhard, G., Jr. (2010). Performance of students with and without disabilities under modified conditions. *Journal of Special Education, 44*, 79–93.

This study evaluated changes in test performance associated with a read-aloud accommodation for students in Georgia in Grades 4 and 7 taking a state reading assessment. Schools were randomly assigned to two conditions: read-aloud delivered via proctor or standard administration. students

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

with disabilities ($N = 459$ in Grade 4, $N = 428$ in Grade 7, mostly learning disabilities, and in Grade 4 also a large percentage of students with speech-language impairments) and students without disabilities ($N = 486$ in Grade 4; $N = 567$ in Grade 7) took the operational exam in the previous year without the read-aloud accommodation and were administered the same test in a low-stakes setting (with or without the read-aloud accommodation) 1 year later. Repeated-measures ANOVA revealed a significant interaction between disability and accommodation use for students in fourth and seventh grades. However, while performance gains for students who received the read-aloud accommodation were significantly higher for students with disabilities (effect size = .22) than for students without disabilities (effect size = .02) in fourth grade, in seventh grade, both student groups received a performance boost from the read-aloud accommodation (effect size = .17 for students with disabilities; effect size = .20 for students without disabilities).

(42) Jackson, R. M. (2012, January). *Audio-supported reading for students who are blind or visually impaired*. Paper prepared for the National Center on Accessible Instructional Materials. Retrieved from http://aim.cast.org/sites/aim.cast.org/files/Audio_Supported_Reading3.8.12_0.pdf

This paper is not a research study but rather a scholarly article that discusses read-aloud for students who are blind or visually impaired, mostly in the context of learning, but it is also applicable to assessment. The authors define audio-supported reading as a combination of refreshable braille or screen magnification and text-to-speech screen reader technology. The following are key points from the article:

- Sensory limitations from blindness or visual impairment are likely directly related to low reading rates; read-aloud is necessary to increase access to information to compensate for slower reading and because braille and large-print materials are not always available when needed in the learning environment.
- Supplementation with speech is therefore considered a necessary tool for increasing access to information not only to compensate for depressed reading rates but also because braille and large-print materials have not always been available when needed.
- With audio-supported reading, reading and comprehension can occur with more efficiency, improving opportunity to learn and acquisition of content knowledge.
- Increased reading rates and information processing speed allow the reader to use working memory to full capacity to comprehend meaning, improve reading comprehension, and decrease the time it takes a student to complete academic tasks.
- Audio-supported reading provides choice and flexibility for the student. When reading passages are more demanding, it is likely that the student will rely more on his or her primary mode (braille or print), and when reading passages are less demanding, read-aloud technology would allow the student to get meaning from the text more quickly.
- There are still challenges for integrating audio-supported reading into classroom practice that relate to teachers of students who are blind or visually impaired, including lack of competency in special assistive technologies, lack of preparation in teaching reading, and many students who are not currently using technology.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

- The materials used to make decisions by IEP teams about the primary mode of learning for literacy instruction should be updated with evidence about the benefits of audio-supported reading for individual students.

Audio-supported reading supports universal design for learning principles by helping students to rely less on aides and more on portable technology.

Table 2: Annotated Bibliography for Audio Presentation for ELA

Study no.	Purpose				Accommodations				Grade			Design			Audio presentation for ELA		
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Read-aloud—text to speech	Read-aloud—pre-recorded Audio/video	Read-aloud—human	Read-aloud—type not specified	Extra time	Elementary	Middle	High	Experimental/quasi-experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference
25	✓	✓				✓	✓		✓				✓		661	801	Different test structure (increased item difficulty, more DIF items) for RLD with and without read-aloud and SWoD, low performing and no accommodations
26		✓				✓	✓		✓	✓	✓		✓		3,900 ^a	6,750 ^a	Different test structure, more so for earlier grades, increased DIF items among SWD with read-aloud, SWD with no accommodations, and SWoD with no accommodations
27		✓					✓		✓				✓		500	500	Similar factor structure for RLD with read-aloud vs. SWoD with no accommodations, SWD with no accommodations, and SWD with accommodations other than read-aloud
28	✓			✓	✓				✓				✓		76	173	Scores higher with read-aloud for SWD and SWoD, differentially higher for SWD
29a,b	✓	✓	✓		✓					✓			✓		237	242	Similar alpha, read-aloud provided little benefit over and above item modification in terms of item difficulty and score gains for SWD eligible for a modified assessment, SWoD, and SWD not eligible for a modified assessment

Study no.	Purpose				Accommodations					Grade			Design			Audio presentation for ELA		
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Read-aloud—text to speech	Read-aloud—prerecorded Audio/video	Read-aloud—human	Read-aloud—type not specified	Extra time	Elementary	Middle	High	Experimental/quasi-experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference	Major findings
30	✓					✓		✓		✓			✓			168	191	Scores higher with read-aloud for RLD and SWoD, no evidence of differential boost, but read-aloud effects were likely confounded with extra time
31	✓					✓		✓		✓			✓			91	91	Scores higher for students with dyslexia and SWoD with read-aloud, differentially higher for SWD
32	✓				✓	✓				✓	✓	✓		✓		168 ^a	3,175 ^a	Higher scores on PPT relative to computer-based test for SWD eligible for a read-aloud accommodation; students generally favored the computer-based test
33	✓		✓			✓	✓				✓			✓		285 ^a	1,643 ^a	Same factor structure and scores for the two different modes of read-aloud for SWD
34	✓		✓			✓	✓					✓		✓		822	44,240	Similar factor structure across SWD who received a read-aloud accommodation, SWoD with no accommodation, and SWD who did not receive a read-aloud accommodation; similar test scores for SWD with and without read-aloud; there was no evaluation of score boost for SWoD
35	✓					✓				✓			✓			15	17	No evidence of score increase associated with read-aloud for SWD or SWoD, but sample sizes were small

Study no.	Purpose				Accommodations				Grade			Design			Audio presentation for ELA			
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Read-aloud—text to speech	Read-aloud—prerecorded Audio/video	Read-aloud—human	Read-aloud—type not specified	Extra time	Elementary	Middle	High	Experimental/quasi-experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference	Major findings
36a,b,c	✓		✓		✓					✓	✓		✓			302 ^a	563 ^a	Evidence of differential boost for RLD relative to SWoD, greater benefit of read-aloud for fourth grade than eighth grade; similar factor structure but more DIF items found (easier) with read-aloud, particularly for fourth grade
37	✓				✓						✓		✓			40	39	No evidence of score differences for read-aloud plus teacher recommended vs. teacher recommended for RLD and SWoD, but sample sizes were small
38	✓					✓					✓		✓			62	198	Scores higher with read-aloud for both RLD and SWoD; not differentially higher for RLD
39	✓		✓			✓				✓			✓			150	145	Different factor structure; test scores significantly higher with read-aloud for both SWD and SWoD; differentially higher for SWD
40	✓		✓		✓						✓		✓			206	200	Similar factor structure; no difference in test scores for SWD and SWoD
41	✓					✓				✓	✓		✓			444 ^a	527 ^a	Scores higher with read-aloud for both SWD and SWoD; differentially higher for SWD in fourth grade, small boost for fourth grade SWoD, moderate boost for both groups in eighth grade

Study no.	Purpose				Accommodations				Grade			Design			Audio presentation for ELA			
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Read-aloud—text to speech	Read-aloud—prerecorded Audio/video	Read-aloud—human	Read-aloud—type not specified	Extra time	Elementary	Middle	High	Experimental/quasi-experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference	Major findings
42				✓	✓					✓	✓	✓			✓	N/A	N/A	This paper is not a research study but rather a scholarly article that discusses audio-supported reading, defined as a combination of refreshable braille or screen magnification and text-to-speech screen reader technology, for students who are blind or visually impaired, mostly in the context of learning, but it is also applicable to assessment

Note. RLD = students with reading-based learning disabilities; LD = students with learning disabilities; SWD = students with any type of disability or unspecified subtypes; SWoD = students without disabilities.

^aStudy has more than one focal and/or reference group, within or across grades. Sample sizes are averaged.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Refreshable Braille for ELA and Mathematics

Refreshable braille displays are hardware that allow digital text to be translated into braille. Many different types of refreshable braille displays exist. The two major differences in products are the number of braille cells available for display at the same time and the type of connections (wireless, USB connection, or connected to a braille notetaker). The use of a refreshable braille display does not appear to alter any of the constructs in the Smarter Balanced Content Specifications. However, the implementation considerations are substantial given the relative novelty of the technology in some states and the diversity of hardware used by students. The following summary provides limited guidance for decision making owing to the lack of research studies conducted in this area.

Summary of Research on Refreshable Braille Displays

Research on refreshable braille displays during assessment is obviously confined to students with visual disabilities. Only three studies were reviewed, and all three contained small sample sizes ranging from 6 to 27 students. Two of the studies were pilot tests of computer-based testing platforms developed to incorporate refreshable braille capabilities. The first study examined a platform developed by AIR for use on the Oregon Assessment of Knowledge and Skills (OAKS) in both mathematics and ELA. The second study examined a platform developed by Measured Progress as part of a U.S. Department of Education grant awarded to ETS on middle school reading assessments delivered via refreshable braille. Finally, a third study examined the viability of using refreshable braille displays on a testing platform (NWEA's Measure of Academic Progress) that was not specifically designed for this purpose. Owing to the sample sizes and the lack of repeated-measures designs, no comparisons can be made between refreshable braille displays and paper-based braille test forms.

Policy implications.

This accommodation (or on-demand braille embossing) is an essential accessibility feature if students are going to participate in the adaptive nature of the Smarter Balanced assessment. However, none of the studies included large enough sample sizes to examine comparability of scores or changes to item difficulty. Important considerations for policy decisions would be the following: (a) Does this feature introduce construct-irrelevant variance for students who only read hard-copy braille? (b) Does Smarter Balanced consider reading refreshable braille a career and college readiness skill? and (c) At what grade level would students be expected to transition from hard-copy braille to refreshable braille?

Implementation considerations.

Several findings emerged regarding implementation considerations. One key finding was that not all test items were amenable to refreshable braille, particularly in mathematics. This will have implications for a computer-adaptive assessment if the features that prevent an item from being brailleable are directly related to content or difficulty, both of which are typical components of item selection in computer adaptive testing. Exclusion of these items may lead to suboptimal proficiency estimates or an inability to cover the required content. Another finding was that rendering items in contracted braille is preferable to uncontracted braille; however, some students expressed the desire to switch back and forth. A third finding was that the user interface for navigating braille content should not override or conflict with common usage of keyboard controls present on

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

refreshable braille devices. In addition, students should have the opportunity to practice using the system prior to testing. Finally, findings point to the need for using accessibility guidelines and best practices such as W3C's Web Content Accessibility Guidelines when utilizing HTML for assessment delivery.

Annotated Bibliography for Refreshable Braille for ELA and Mathematics

(43) Katz, M. (2012). *Computer-based testing accommodations for students with visual disabilities: Alpha test*. Unpublished manuscript.

This study was a preliminary field trial (“alpha test”) of a refreshable braille capability for the NimbleTools computer-based assessment platform to identify potential problems prior to conducting additional development. Ongoing research will include two additional iterative studies followed by a larger field trial. This preliminary study was designed to examine the basic usability of the refreshable braille interface and utilized a single reading passage (eighth-grade level) previously developed for the DARA project. Students ($N = 6$) were recruited from Grades 8 to 12 via teachers of students with visual impairments (TVIs) in Massachusetts, and the study was carried out in the students’ schools. A researcher and a TVI were present during the sessions, which lasted approximately 1 hour. To focus specifically on the braille interface, the audio presentation feature within NimbleTools was disabled. The session included direct observation and a 15-question cognitive interview. Though all six students were able to complete the reading passage and related test items, and responded positively to the system, a number of issues in the design of the interface were identified. These issues included students’ preference for Grade 2 (contracted) braille (in contrast to the Grade 1 uncontracted braille provided by the alpha test system), formatting of information presented on the braille display, and the use of navigation keys on the braille device.

(44) Kamei-Hannan, C. (2008). Examining the accessibility of a computerized adapted test using

This study examined the accessibility of the Measure of Academic Progress (MAP), a computerized adaptive test, by students with visual impairments who required either braille or magnification. The study participants were students at the Arizona School for the Deaf and Blind. Students whose primary literacy mode was braille were administered reading and language tests using a refreshable braille display (32 or 80 cells) driven by the JAWS 5.1 screen reader. Though MAP included a math component, the braille users were not administered this portion of the assessment owing to the inability of JAWS to render Nemeth code. Students utilizing the braille display were required to have skills in document navigation and the placement and routing of the cursor. For the reading assessment, the braille users ($N = 27$) found that 20.5% of the 42 items administered were not answerable owing to accessibility issues, including scrolling of long passages, inconsistent or undetectable underlining, spatial formatting, graphical elements and pictures, and braille translation errors. In the language assessment, the braille users ($N = 28$) found that 13.12% of the 52 items on the assessment were not answerable owing to accessibility issues. In the language assessment, the most common problem was underlining and included spatial formatting, graphical elements, and braille translation issues. The author discusses limitations in the study, including student experience with braille, the ability of the proctor to recognize that a student was encountering accessibility issues with an item, and the probable improvements in the assistive technology (JAWS) over time. Because the assessments were delivered using HTML, improvements in the application of accessibility guidelines may mitigate some of the accessibility issues encountered.

(45) Oregon Department of Education. (2011). *Braille interface of OAKS Online—May 2011 proof of concept*. Retrieved from <http://www.ode.state.or.us/search/page/?id=3369>

This study describes a pilot test to evaluate the braille interface of the OAKS. The study specifically access, navigate through, and enter responses on the test. Additionally, the study examined the students’ comfort using the supporting Literature Review of Testing Accommodations and Accessibility

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

braille technologies used to access the test, the functioning of the supporting braille technologies in an assessment setting, and training needs for both students and teachers of students with visual impairments. Students ($N = 11$) from Grades 3, 5, 6, 7, and 8 and from high school, with a range of braille skills, were presented with an 8-item practice test in one of the content areas of reading, math, science, or social sciences. The practice test included both text-based items and items that included a spatial component such as a table or graph, representing a sample of the items the student would encounter on the operational assessment. An interview was conducted with the student after completion of the practice test. Results from the interviews indicate that students were comfortable with test format and braille technologies, which included both a refreshable braille display and a tactile embosser. Some students commented that the online test seemed to go faster and was more fun than the paper-based braille test. A survey of students' teachers of the visually impaired was also conducted. The teachers reported that the students participating in the pilot test had at least some skill with using computer keyboards and reading embossed braille materials, and almost all students had at least some skill with using a screen reader and refreshable braille display. Both the interviews and survey indicate that the students were comfortable using the braille technology and that the system was easy to learn. However, the study also indicates that feedback suggests additional training is recommended on the online braille system for students, TVIs, and test administrators.

Table 3: Annotated Bibliography for Refreshable Braille for ELA and Mathematics

Study no.	Purpose				Grade			Design			Refreshable braille for ELA and mathematics			
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Elementary	Middle	High	Experimental/quasi-experimental	Operational data	Qualitative/descriptive	Test content	Sample size focal	Sample size reference	Major findings
43				✓		✓	✓			✓	ELA	6	N/A	All students with visual impairments completed the passage and test items and responded positively to the accommodations, but there were some implementation issues
44				✓	✓	✓	✓			✓	ELA	28	N/A	13% and 20% of items were found to be inaccessible for students with visual impairments
45				✓	✓	✓	✓			✓	ELA/math	11	N/A	Students with visual impairments and teachers indicated that students were comfortable using the braille technology and that the system was easy to learn; additional training was recommended

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

American Sign Language (ASL) for Mathematics

Hearing disability (deaf/hard of hearing) does not preclude the ability to read math symbols or English words. Yet difficulties in English language literacy often associated with hearing disability may pose accessibility challenges. For some students with hearing disabilities, English is a second language, with ASL being the first language. On the basis of a review of the Smarter Balanced Mathematics [Content Specification](#), there does not appear to be an explicit requirement that students must read and comprehend the English language. For this reason, the translation of text into ASL may be an appropriate accommodation. The delivery of ASL in mathematics can take the form of prerecorded video, avatars, or a teacher signing. In the case of teacher signing, this can occur by following a detailed script, following general guidelines, or completely on the fly. In addition, the content provided in ASL may vary from isolated words and phrases to the entire assessment. The following summary provides limited guidance for policy makers on the use of ASL as an accommodation.

Summary of Research on ASL for Mathematics

This review included 11 studies that examined the impact of sign translation in an assessment context. Three experimental or quasi-experimental studies examined the relationship between ASL interpretation of mathematics items and test performance by students who are deaf or hard of hearing. Of these studies, two found no significant differences in performance, and the third did not allow for comparisons owing to the lack of equated forms. In addition, one study examined DIF between examinees who did not have any accommodations and examinees with any of a range of disabilities who received a presentation accommodation (e.g., interpret items using sign language or any of several other presentation accommodations) and found five items that were easier for the accommodated group and four that were easier for the nonaccommodated group. Owing to considerations such as lack of detailed breakout of data for ASL item interpretation, few inferences about ASL item interpretation are possible.

Several qualitative studies examined ASL translation issues for math items and found that some item types were more difficult to translate (particularly those that involved graphical or tabular material) and may have altered item difficulty. These changes in item difficulty have potential implications for the comparability of scores and selection of test items in a computer-adaptive assessment. Generally, it was found that there is little research available on the influence of ASL in the presentation of mathematics items. In addition, the limited research indicates that a number of factors can prevent ASL presentation of items from having the desired impacts. Among these are limitations in students' familiarity with ASL, errors in translation from English to ASL, and limitations in students' math knowledge (which can prevent removal of accessibility barriers from improving math scores).

Another area of research included studies that examined professional decision making and policy trends in accommodations for students who are deaf or hard of hearing. Cawthon (2011a) conducted a quasi-experimental study that examined education professionals' decision making regarding accommodations for students who are deaf or hard of hearing. She concluded that a more systematic approach is needed to make valid inferences about scores obtained when using accommodations. The two other studies involved surveys that shed light on current educational practices and other issues affecting the provision of ASL translations in mathematics. These studies

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

found that 24% of teachers reported signing math test items, 35% reported signing math and reading test content, and 27% of teachers reported allowing students to sign responses.

Finally, two other studies examined how students approach items in ASL and had interesting findings but were not directly related to accommodation policy or implementation. For example, a qualitative study examined how students in grades K–3 responded to ASL items presented via videotape. The study found that half of the children used “viable” strategies (i.e., that could result in a correct answer) for at least half the items. Another study (Cawthon, 2011b) used content analysis to analyze Grade 5 and Grade 8 reading and math items and then had students in Grades 5–8 respond to the items in regular paper-and-pencil format. She found that the average linguistic complexity score of math items was significantly higher than for the reading items.

Policy implications.

Given the small sample sizes, there is little information on the psychometric comparability of scores. However, some of the qualitative studies demonstrated changes in item difficulty so it would be advisable to monitor item difficulty and DIF to determine if changes in item difficulty would impact either measurement of growth or the adaptive nature of the Smarter Balanced assessment.

Implementation considerations.

Though the quantitative research is limited, several of the studies provided rich data that could inform implementation considerations. First, it is critical to involve expert signers (native signers if feasible) and educators of students who are deaf or hard of hearing during item development to ensure that test items can be translated into ASL without changing the construct or significantly altering difficulty. This is particularly important for ASL translations of graphical and tabular material and geometrical concepts (parallel, perimeter). Second, it is clear that additional training on assigning accommodations would be helpful to ensure that both test content and student’s instructional experience with an accommodation are consistent. Finally, it may be worthwhile to consider use of animated ASL as well as recorded human ASL and avoid use of synchronized highlighting of text (or other simultaneous visual representations) while signing is occurring.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Annotated Bibliography for ASL for Mathematics and Reading

(46) Ansell, E., & Pagliaro, C. (2001). Effects of a signed translation on the types and difficulty of arithmetic story problems. *Focus on Learning Problems in Mathematics*, 23(2-3), 41-69.

This qualitative study described and evaluated the changes in item type and difficulty that occurred when math story problems were translated into “sign language.” Subjects ($N = 38$) were more than 90% of the primary-level (K-3) teachers in schools for the deaf in five states. Seven teachers were deaf, 3 were hard of hearing, and 27 were hearing. The native language of teachers varied: 6 teachers reported themselves to be native ASL signers, 2 listed themselves as native to signed English, and 29 were native speakers of English. Each teacher was videotaped signing each of 15 story problems “as they would for their students” (p. 55). Teachers were allowed to practice and could re-sign a problem before going on to the next problem. The researchers then coded and analyzed the version of the problem with which the teacher expressed satisfaction. The videotaped recordings were then translated into written English by one of the researchers. The signed versions and the English translation were then coded according to the attributes specific to problem types (e.g., the inclusion of situation action) and aspects related to problem difficulty (e.g., the change of chronology of events; p. 55). The researchers found that of the 15 story problems translated from written English to sign language, 3 reflected a change in problem type, and 10 exhibited a shift in difficulty within problem type. Sign language variations used within problem type as well as the dynamic, visual nature of sign language can clarify the relationships and/or actions depicted within a problem situation. However, these may lessen the inference needed by the student to move from problem statement to its modeled solution and therefore decrease problem difficulty. Furthermore, the results indicate that both the presentational features of sign language and teachers’ choices can create either bridges or barriers to deaf students’ understanding. Changes in problem type resulting from the interaction of sign language features and teachers’ choices may limit the types of problems deaf and hard-of-hearing students are given to solve, thereby restricting their access to mathematical understanding. The findings of this study raise important issues about the translation of arithmetic word problems into sign language. This research should cause professionals in deaf education to reflect on the mathematics they present to their students. Too often, teachers and teacher educators in the field of deaf education focus on linguistic issues without considering the resulting effects on content knowledge. For example, “it is imperative that defining characteristics of those problems be preserved in translation. The translator (e.g. teacher, interpreter) must be aware of these characteristics and their implications, as well as how specific sign features affect them” (p. 66).

(47) Ansell, E., & Pagliaro, C. (2006). The relative difficulty of signed arithmetic story problems for primary level deaf and hard-of-hearing students. *Journal of Deaf Studies and Deaf Education*, 11(2), 153-170.

This study examined the relative difficulty and associated strategy use of arithmetic (addition and subtraction) story problems in math when presented in ASL (on videotape) to primary-level (K-3) students who are deaf or hard of hearing ($N = 233$). The six story problems were selected to be comparable to those used in studies involving hearing children. To develop the ASL recordings, the researchers involved four deaf native signers of ASL to determine the translation from written English to ASL. The ASL versions were recorded on videotape. The problems were signed to be appropriate for primary-level children, follow the rules of ASL, and maintain the mathematical structure of the original problems. The results showed that half of the children used viable strategies

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

on more than half of the six problems presented. Additionally, students who are deaf or hard of hearing may consider and respond to arithmetic story problems differently than their hearing peers, with the key dimension in problem difficulty being based on the operation typically used to solve the problem (as opposed to the story within the problem). Most children did not appear to view the signing of the problem as containing links to the solution; many children “did not seem to attend to the problem situation at all, focusing primarily on the numbers in the problems. They ignored or did not recognize any relationship between the story and its solution, thus missing linguistic markers that could potentially have made for an easier problem” (p. 167). Thus the visual-spatial nature of the ASL problem presentation did not appear to assist the students in solving the problems.

(48) Cawthon, S. (2011a). Making decisions about assessment practices for students who are deaf or hard of hearing. *Remedial and Special Education, 32*(1), 4–21.

This study examined teacher recommendations for assessment practices for students who are deaf or hard of hearing. Participants were 372 educational professionals—teachers and other education professionals who worked with students who were deaf or hard of hearing in a range of settings (Grades K–12). Each participant responded to a set of three study vignettes that asked for recommendations for accommodations or alternate assessments appropriate for the scenarios. Four randomly assigned conditions controlled for test subjects (math or reading), student skill level (two or five grades below grade level in math and/or reading), and communication mode used in instruction (ASL or Total Communication [a combination of sign and speech]). Findings revealed that the most common accommodations recommended were interpretation of test directions, extra time, and interpretation of test items. Findings also indicated that recommendations for accommodations differed by subject (math vs. reading) and by student proficiency (high vs. low), but communication mode (ASL vs. Total Communication) was not a significant factor in choosing accommodations. The author questioned whether score validity was being adequately considered (pp. 15–16). Furthermore, “there was not one predominant variable across all scenarios as one might expect to find if participants were taking a systematic approach to accommodations decision making” (p. 18). She further suggested that “one could build a flow chart model of criteria in decision making, with policy for use as an early filter in the process” (p. 19).

(49) Cawthon, S. (2011b). Test item linguistic complexity and assessments. *American Annals of the Deaf, 156*(3), 255–269.

This exploratory study examined the relationship between linguistic complexity and test performance for deaf readers. Subjects were 64 students in Grades 5–8 (ages 10–15 years) in schools for deaf students. Each student was administered 52 multiple-choice items (released items from Grade 5 and Grade 8 Texas state assessments). Math items were word problems, and reading items had a reading passage. Items were coded for their linguistic complexity. The researchers found that there were small relationships between an item’s linguistic complexity and item difficulty; the strongest relationship is for the syntax rating for mathematics items, with a negative correlation at $r = -.22$; thus the higher the syntax rating, the less likely students were to answer the item correctly. Though reading is typically thought of as being more linguistically demanding than math, the results showed that the average linguistic complexity score of math items was significantly higher than for the reading items ($p < .05$). The researchers hypothesize that “a student’s knowledge of how to interpret a mathematics word problem resided primarily within the student (and his or previous experiences with the content)” (pp. 262–263). Conversely, for reading items, students were able to refer to a reading passage, which served as “an identifiable, presented knowledge base” to help students infer meanings (p. 262). One recommendation from this study was to create rules that allow for the use of

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

ASL to reduce the linguistic complexity of math items without violating the construct being measured. Rules may vary depending on the nature and part of the math item (e.g., directions, graphics, math expressions [equations], and other test content).

- (50) Cawthon, S. W., & Online Research Lab. (2009). Accommodations for students who are deaf or hard of hearing in large-scale, standardized assessments: Surveying the landscape and charting a new direction. *Educational Measurement: Issues and Practice*, 28(2), 41–49.

This study presents findings from the Third Annual Survey of Assessments and Accommodations for students who are deaf or hard of hearing, which was conducted in the 2006–2007 school year and involved educational professionals (teachers of students who were deaf or hard of hearing in schools for the deaf, district/regional programs, and mainstreamed classrooms; $N = 389$) involved with students who are deaf or hard of hearing in a variety of settings in Grades K–12. Among the findings were that accommodations of small group, extended time, and test directions interpreted were used extensively for both math and reading assessments. A large proportion (73%) of respondents did not use the “student signs response” accommodation. Also, “across the board participants rarely gave an accommodation for reading that was not also given for math,” this being a “far more pronounced effect than from previous results” (p. 45). For prevalence of test items interpreted, the rates were 24% for math only and 36% for both math and reading. Among the communication modes (ASL, other signed language, oral [speech], Total Communication [oral and sign together by instructor], oral [speech] plus interpreter, other), the most prevalent mode for each educational setting was ASL in schools for the deaf (78% of participants in that setting), oral and sign language together in district/regional programs (72%), and oral only in mainstreamed programs (69%), respectively. The percentage of participants indicating use of the ASL mode in their setting was 78% for schools for the deaf (as mentioned earlier), 42% for district/regional programs, and 20% for mainstreamed programs (p. 44).

- (51) Cawthon, S. W., Winton, S. M., Garberoglio, C. L., & Gobble, M. E. (2011). The effects of American Sign Language as an assessment accommodation for students who are deaf or hard of hearing. *Journal of Deaf Studies and Deaf Education*, 16(2), 198–211.

This quasi-experimental study examined the effects of presenting items in ASL in reading and math to Grade 5–8 students ($N = 64$; ages 10–15 years) in schools for the deaf where ASL was the primary mode of instruction. The recordings were created by two native ASL signers, both fluent in ASL and English. The study was administered by displaying a video recording of a human signer either on individualized computer screens or on a large screen (via DVD). In the ASL condition, students responded via printed items in a booklet. The researchers found that there were no overall differences in the mean percentage of items students scored correctly in the standard versus ASL-accommodated conditions for reading or math. A hierarchical linear regression was conducted to determine whether measures of exposure to ASL (home and classroom) and student proficiency in the subject area predicted student performance in ASL-accommodated assessments. The models predicted up to half of the variance in the scores, with subject area proficiency (mathematics or reading) as the strongest predictor. ASL exposure was not significant, except for ASL classroom instruction as a predictor of mathematics scores.

- (52) Johnson, E. S., Kimball, K., & Brown, S. O. (2001). American Sign Language as an accommodation during standards-based assessments. *Assessment for Effective Intervention*, 26(2), 39–47.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

The researchers examined political, practical, and psychometric issues related to the use of sign language (ASL and Signing Exact English [SEE II]) for math and listening in items derived from the Washington State Assessment of Student Learning; this review focuses mainly on math, specifically, the adequacy of the ASL translations of math items for deaf and hard of hearing students in Grades 4, 7, and 10. (The number of students involved was not disclosed.) At each grade level, 20 of the 40+ items (40, 46, and 40 at the respective grades) were interpreted by certified ASL signers and recorded on videotape. Other certified interpreters then viewed the videotapes and translated the items back into oral or written English. The transcriptions were then analyzed to determine whether a student with the sign language version of the item would “(a) obtain the information required to complete the items correctly, (b) receive an unfair advantage owing to the conceptual nature of sign language, or (c) be misled by possible erroneous information in the signed version. Two independent reviewers analyzed each set of transcriptions and the results were compared. A third independent reviewer resolved discrepancies” (p. 41). At the Grade 4 level, the analysis suggested that problems in translation were most common with graphical or tabular material. Specifically, seven of the eight “unanswerable” items (item a) involved a bar graph, table, or number line that related to questions. In all seven of these cases, “information depicted in the graph was misinterpreted in a way that would be misleading to the student” (p. 42). Furthermore, five of the eight unanswerable items were open ended as opposed to multiple choice. The patterns were not as clear at Grades 7 and 10; however, at those grades, there were also similar numbers of items that were judged as unanswerable. The researchers gave an example of an ASL translation issue that underscores the importance of detailed definitions of the construct to be measured. They give the example of the need to determine what to do when the use of a sign changes the nature of what is assessed, for example, the requirement for knowledge about the meaning of the term *perimeter* is reduced when the sign indicates that it is the distance around a figure. The researchers offer several suggestions for practice: (a) Provide training for interpreters to “standardize the use of any approved sign language, especially ASL” (p. 46); (b) recognize the complexity of content and assessment format—interpreters must have sufficient time and content knowledge (e.g., science, math); (c) anticipate time frame and schedule changes—extra time is needed to administer tests to students who are deaf or hard of hearing students; and (d) monitor use (and abuse) of accommodations.

(53) Maihoff, N. A., Bosso, E., Zhang, L., Fischgrund, J., Schulz, J., Carlson, J., et al. (2000). *The effects of administering an ASL signed standardized test via DVD player/television and by paper-and-pencil: A pilot study*. Dover: Delaware Department of Education. Retrieved from <http://www.doe.k12.de.us/aab/files/Atch9.pdf>

This study examined the effects of administering an ASL signed standardized math test via DVD to deaf students ($N = 19$; ages 10–11 years) in two schools for deaf students. The researchers administered two tests in group administration: (a) one composed of multiple choice items from the CAT6 standardized math test via DVD with an accompanying paper-based test and (b) the Terra Nova test in its regular written format. The study had been intended to test the null hypothesis that “a standardized test that is signed and presented in video format (CAT6) would show the same performance” (p. 2), as on the Terra Nova form. Although “both the CAT6 and Terra Nova are on the same scale” (p. 3), the tests had not been equated so that “without the meaningful equated standard scores for both tests, the hypothesis could not be tested” (p. 4). The report indicates that “several students” reported “that they preferred a written test over the DVD version because the DVD version was too slow; however they did report that the signed test was easier to understand, and that the ASL signing was clear” (p. 5). “Overall the test administrators felt that the signing of the

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

test on DVD was more clear” and a “much better presentation than having individual teachers sign to students” (p. 5).

(54) Qi, S., & Mitchell, R. E. (2012). Large-scale academic achievement testing of deaf and hard-of-hearing students: Past, present and future. *Journal of Deaf Studies and Deaf Education*, 17(1), 1–18.

This report offers insights and suggestions for testing programs designed to assess what students who are deaf or hard of hearing know. It summarizes historical data over the last 3 decades to indicate trends in academic achievement for this special population, analyzes the current federal laws and regulations related to educational testing and special education, and identifies limitations of current testing programs in assessing what deaf students know and can do. The study noted that

testing in the English language may inhibit the ability of deaf and hard-of-hearing students to fully express what they know. However, the potential solution of translating standardized tests into ASL has not been widely adopted for large-scale testing due to lack of psychometric studies . . . , as well as practical considerations, such as funding. (p. 3)

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

The authors went on to write that

Translating the test into ASL would, in principle, be appropriate for certain types of mathematical questions, but not for a test of English reading comprehension. In the latter case, written English is a central feature of the construct being measured by the instrument. (p. 11)

However, use of ASL even for reading and language arts might be appropriate if the student were classified as a “limited English proficient student” where a deaf student communicates primarily in ASL. “Only South Carolina has systematically undertaken to provide statewide assessments in ASL” (p. 14). The authors noted that “translation is likely to benefit students who learned their subjects in ASL” (p. 13), that “translation is not a simple matter of word-by-word translation” (p. 13), and that “validity studies are required to determine the extent to which any test performance difference [between the ASL and the written English versions] is explained by the presentation change associated with ASL adaptation” (p. 13). They indicated that “extensive psychometric research employing experimental designs is urgently needed for future test development in this area” (p. 15).

(55) Russell, M., Kavanaugh, M., Masters, J., Higgins, J., & Hoffmann, T. (2009). Computer-based signing accommodations: Comparing a recorded human with an avatar. *Journal of Applied Testing Technology*, 10(3). Retrieved from http://www.testpublishers.org/assets/documents/computer_based.pdf

This study employed a quasi-experimental design to examine the impact on math performance of the manner of presenting math test content in a computer-based ASL format—specifically, with a human signer (via video) versus an avatar signer (via animation). The study also examined student preference for format. Ninety-six Grade 8–12 students who were in schools for students who are deaf or hard of hearing were administered two parallel forms of a Grade 8 math test (using released multiple-choice NAEP math items). The study involved a multistep process for producing the two forms. First, a script was created that described how the item was to be presented in ASL. Second, a human signer was recorded for each item. Third, the recording was reviewed by experts, and suggestions for modification were made. Fourth, the human signer was rerecorded; this video was the recorded human signer form. Next the recording of the human was used by a company specializing in ASL animations (Vcom3d) to produce the animations used in the avatar signer form. The two forms—human and avatar—were then implemented in a delivery system that allowed the student to see the appropriate version and respond to the item; the ordinary English text version of the test item was also visible on the screen simultaneously. Students were assigned to groups that determined the order of forms, one of which had the human signer and the other the avatar signer. The researchers found that there were no significant differences in math performance between the two formats of ASL accommodation, and this finding held at varying performance levels. Additionally, about two-thirds of the participants expressed a preference for the human signer, whereas one-third preferred the avatar signer. The use of recorded human versus animated avatar did not affect the amount of time required to complete the test items. Students expressed positive reactions to computer-based test delivery and the accommodations.

Table 4: Annotated Bibliography for ASL for ELA and Mathematics

Study no.	Purpose				Accommodations		Grade			Design			ASL for ELA and mathematics			
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Avatar signer (computer animation)	Human signer (via video)	Elementary	Middle	High	Experimental/quasi-Experimental	Operational data	Qualitative/descriptive	Test content	Sample size focal	Sample size reference	Major findings
46	✓		✓		✓		✓					✓	Math	38	N/A	Teacher signing of mathematics problems results in changes to difficulty (10–15 items) and in some cases changes to the problem type (3 of 15 items)
47	✓		✓		✓		✓					✓	Math	233	N/A	Students who are deaf or hard of hearing respond to “story problems” differently than their hearing peers; when stories are signed, the primary contributor to item difficulty is the mathematical operations and not the story within the problem
48			✓				✓	✓	✓			✓	ELA/ Math	372	N/A	Teachers of students who are deaf or hard of hearing are most likely to assign interpretation of test directions, extra time, and interpretation of test items as accommodations when presented with vignettes of students; decisions changed by content area and proficiency level but not communication mode of student
49			✓				✓	✓				✓	ELA/ Math	64	N/A	Small negative relationship between linguistic complexity of items and performance for students who are deaf

ASL for ELA and mathematics																	
Study no.	Purpose				Accommodations		Grade			Design				Test content	Sample size focal	Sample size reference	Major findings
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Avatar signer (computer animation)	Human signer (via video)	Elementary	Middle	High	Experimental/quasi-Experimental	Operational data	Qualitative/descriptive					
50				✓			✓	✓	✓			✓	ELA/ Math	389	N/A	According to teachers in schools for the deaf and mainstream classrooms, small group, extended time, and test directions interpreted were used extensively; ASL was used most in schools for the deaf	
51	✓				✓		✓	✓		✓			E*LA/ Math	64	N/A	No evidence of score differences between ASL and standard administration for student in schools for the deaf, where ASL was the primary mode of instruction but the sample size was small	
52				✓	✓		✓	✓	✓			✓	ELA/ Math	undisclosed	N/A	Some math item types (graphic and tabular) are more difficult to translate into ASL and may result in “unanswerable” items; several recommendations are provided on how to translate the findings to practice in large-scale assessment	
53	✓				✓		✓			✓			Math	19	N/A	Scores from students who are deaf taking both a standard form and with ASL could not be compared because the accommodated and standard assessments were not equated to be on the same scale	

ASL for ELA and mathematics																
Study no.	Purpose			Accommodations		Grade			Design			Test content	Sample size focal	Sample size reference	Major findings	
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Avatar signer (computer animation)	Human signer (via video)	Elementary	Middle	High	Experimental/quasi-Experimental	Operational data					Qualitative/descriptive
54				✓			✓	✓	✓			✓	ELA/ Math	N/A	N/A	This report suggests that testing in English is a barrier to access for students who are deaf and hard of hearing and that ASL should be considered as an accommodation for mathematics assessments and in limited cases for ELA assessments
55	✓				✓	✓		✓				✓	Math	96	N/A	No evidence of score differences across accommodation modes for students who are deaf or hard of hearing; the majority preferred the human signer

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Calculators for Mathematics

A review of the Smarter Balanced Mathematics Content Specifications indicates that in some early grades, computation fluency is part of the construct. However, this accommodation does not appear to impact other early grade mathematics content, and some higher grade levels require the use of computation tools such as spreadsheets. The use of calculators on an assessment can take different forms that vary in the features of the calculator (e.g., four-function, scientific) and the format (handheld or on-screen). The following information provides guidance for making decisions about calculator use as an accommodation for students with disabilities. Because there were so few studies on calculator use that included students with disabilities, we extended our review to include quantitative studies on students without disabilities only, provided that they were experimental/quasi-experimental and focused on assessing K–12 students (including studies on the SAT®).

Summary of Research on Calculators for Mathematics

The results of the four studies that include students with disabilities suggest that students with disabilities in elementary and middle school do not receive differentially higher score gains from calculator usage on problem-solving assessments. These studies sampled students with disabilities in general or students with learning disabilities, broadly speaking. Two of the four studies had small sample sizes—as such, the finding of no significant interaction cannot be interpreted as evidence of no differential boost for students with disabilities. There may be students with more specific disability subtypes who would benefit differentially from calculator use, but this has not been studied. Two additional studies summarized that include only students without disabilities in high school indicate that the use of a calculator may provide a benefit for complex computational items and may be associated with decreased performance on some item types. These two studies focused on item characteristics, whereas the four studies that included students with disabilities did not.

Policy implications.

Overall, it does not appear that calculator use provides more of a benefit to students with disabilities than to students without disabilities, and in some cases, it is associated with unexpected decreases in scores. However, none of the available research has focused on students with specific disability subtypes related to mathematics such as dyscalculia. A related body of research has focused on learning and skills associated with calculator use in the classroom (with and without calculator use during testing). This research is summarized in a meta-analysis (Ellington, 2003) and suggests some benefits to students in general in terms of learning and skill acquisition in K–12 mathematics classes; however, the findings may not be generalizable to students with disabilities or to low-performing students.

Implementation considerations.

Were calculators to be permitted for use by some or all students, students should be given the opportunity to use the calculator in their classroom activities prior to testing. Otherwise, student agility with calculator usage may become a construct-irrelevant factor measured by the mathematics items. An online calculator is problematic for students with visual impairments; such students would need access to an acceptable personal calculator.

Annotated Bibliography for Calculator for Mathematics

Studies that include students with disabilities:

- (56) Bouck, E. C. (2009). Calculating the value of graphing calculators for seventh-grade students with and without disabilities: A pilot study. *Remedial and Special Education, 30*, 207–215.

This small study explored performance differences associated with using a graphing calculator on a seventh-grade mathematics assessment with problem solving focused on numbers and operations. Students with disabilities ($N = 13$) and students without disabilities ($N = 27$) from two classes in one school district in a midwestern state participated in the study. All students took a pretest with no calculator and a posttest 1 month later. Students in one classroom took the posttest with a graphing calculator, and students in the other classroom took the assessment with no calculator. All students had access to a graphing calculator during class in the time in between the pretest and posttest. Both tests were timed, and many students did not complete the test. The analysis method was ANCOVA with pretest scores as covariate. Classroom and accommodation condition were confounded in this study. Results showed no significant interaction between disability status and accommodation condition, and students who received the calculator scored higher on average, controlling for pretest performance.

- (57) Bouck, E. C., & Bouck, M. K. (2008). Does it add up? Calculators as accommodations for sixth grade students with disabilities. *Journal of Special Education Technology, 23*(2), 17–32.

In this study, sixth-grade students from six classes in three schools in a midwestern state took two versions of a mathematics assessment consisting of problem solving focused on numbers and operations. Each class had at least one student with a disability. Students with disabilities ($N = 22$) and students without disabilities ($N = 67$) each took a pretest and a posttest. Students in half of the classrooms took the posttest with a four-function calculator. The tests were timed, and many students did not finish. In the 6 weeks in between pretest and posttest, students used the calculator in classroom activities. The analysis approach was ANCOVA with pretest scores as covariate. Results showed no significant interaction between disability status and accommodation condition and suggested that students in both groups scored higher, on average, when they took the test with a calculator.

- (58) Engelhard, G., Jr., Fincher, M., & Domaleski, C. S. (2010). Mathematics performance of students with and without disabilities under accommodated conditions using resource guides and calculators on high stakes tests. *Applied Measurement in Education, 24*(1), 22–38.

This study evaluated the association between calculator use and performance on the Georgia Criterion Referenced Test mathematics assessment. A random sample of schools across the state was chosen, and schools were randomly assigned to have students in fourth and seventh grades take the assessment under standard conditions, with a calculator, or with a resource guide. All students took the same state assessment in the previous year in the usual high-stakes setting. The sample included students with disabilities ($N_4 = 459$; $N_7 = 430$) and students without disabilities ($N_4 = 488$; $N_7 = 567$). The number of students with different disability subtypes was close to percentages in the entire state, mainly composed of students with learning disabilities or, in the lower grades, speech and language impairment. The calculator was a basic function calculator, and the students were given the opportunity to practice with it. The analysis method was ANCOVA with the prior year's scores as a covariate. Although there was random assignment to conditions, students in the standard condition scored higher, on average, on the nonaccommodated prior-year

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

test in both groups in Grade 3 and in the group of students with disabilities in Grade 7. Results suggested that in fourth grade, scores were significantly higher with the calculator relative to the standard conditions, after controlling for prior-year performance, for students with disabilities but not for students without disabilities. However, the interaction between condition and disability group was not significant. In Grade 7, there was no evidence of a differential benefit for students with disabilities relative to students without disabilities. Students without disabilities experienced a small average score increase when using the calculator, while average scores for students with disabilities slightly decreased.

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

A subset of this study focused on calculator use. Students with learning disabilities ($N = 192$) and students without disabilities ($N = 181$) in Grades 4 and 5 took concepts and application and problem-solving curriculum-based measurements both with and without a calculator. Administration order was counterbalanced. Read-aloud was offered with the calculator for the problem-solving curriculum-based measurement. ANOVA results showed no significant interaction of condition and disability status, and there was evidence that scores stayed the same or decreased when students were permitted to use calculators.

Studies that do not include students with disabilities (note that only experimental studies are included; observational studies are excluded):

- (60) Bridgeman, B., Harvey, A., & Braswell, J. (1995). Effects of calculator use on scores on a test of mathematical reasoning. *Journal of Educational Measurement, 32*(4), 323–340.

A sample of high school juniors from 275 high schools ($N = 11,457$) who planned to attend college took an experimental version of the SAT mathematical reasoning test comprising mostly previously released operational items. A random half of the sample was permitted to use a calculator; for the other half, calculator use was not allowed. Item types included regular mathematics (multiple-choice and student-produced response) and quantitative comparisons; no items required a calculator. Results indicated that total test score was slightly higher for the calculator group and that item-level differences across groups varied in direction and magnitude. Prior experience with calculator use was associated with increased test performance for students in the calculator group. There was some evidence that students would benefit from the use of a calculator more when there were more items requiring complex computation. Higher performance associated with calculator use was found at all levels of ability and item difficulty, indicating that generalizations about the benefits of calculator use to low- or high-scoring groups could not be made.

- (61) Loyd, B. H. (1991). Mathematics test performance: The effects of item type and calculator use. *Applied Measurement in Education, 4*(1), 11–22.

Students in 8th–11th grades who attended a summer enrichment program took an experimental mathematics assessment. The test included both items that did not require calculator use (sums up to 10 or division by 10) and those developed for calculator use (computation and estimation); additional items included those similar to the Iowa Test of Educational Development, which tested process knowledge and item types similar to those on the GED that were found in prior research to be negatively associated with calculator use. Students were randomly assigned within blocks to the calculator group ($N = 70$) or the noncalculator group ($N = 90$). Results supported the hypothesis that calculator use is more beneficial for complex computation and estimation items. There were also no

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

significant differences in test scores between groups on the items testing process knowledge and some evidence that there are some item types in which a calculator is a hindrance. Calculator use appeared to be unrelated to the amount of time students took to complete the test.

Two additional studies, Scheuneman, Camara, Cascallar, Wendler, and Lawrence (2002) and Walcott and Stickles (2012), are observational, using extant data that did not include students with disabilities. Because differences in test scores are confounded with choice in using calculators, and because students with disabilities are not included, the results do not contribute relevant information on calculators as an accommodation for students with disabilities.

Table 5: Annotated Bibliography for Calculator Mathematics

Study no.	Purpose				Accommodation			Grade			Design			Calculator for mathematics		
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Graphing calculator	Four function calculator	Calculator – type not specified	Elementary	Middle	High	Experimental/quasi-experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference	Major findings
55	✓				✓				✓		✓			13	27	Students who received the calculator scored higher, on average, controlling for pretest performance and there were no differential effects associated with disability status. However, classroom and accommodation condition were confounded
56	✓					✓				✓	✓			22	67	Students who received the calculator scores higher, on average, controlling for pretest scores, and SWD did not receive a differential benefit from the calculator
57	✓					✓			✓	✓	✓			445 ^a	528 ^a	In Grade 4, scores were significantly higher with the calculator for SWD but not for non-disabled students, but the interaction was not significant in the combined analysis. In Grade 7, scores were not significantly higher with the calculator for either group
58	✓						✓				✓			192	181	There was no significant interaction with disability status and calculator use. There was evidence that scores stayed the same or decreased with calculator use for all students, on average

Calculator for mathematics															
Study no.	Purpose				Accommodation			Grade			Design		Sample size focal	Sample size reference	Major findings
	Compare scores	Item comparability	Compare test structure	Descriptive/other	Graphing calculator	Four function calculator	Calculator – type not specified	Elementary	Middle	High	Experimental/quasi-experimental	Operational data			
59	✓					✓				✓	✓		5,729	5,729	*Students without disabilities only. Test scores were slightly higher with calculator use but item-level differences across accommodated groups varied in magnitude and direction. Some evidence that items with complex computations were more likely to be answered correctly with a calculator
60	✓					✓			✓	✓	✓		70	90	*Students without disabilities only. Calculator use was associated with higher scores for complex computation and estimation items. There were no significant score differences associated with calculator use for items testing process knowledge and there were some item types in which the calculator introduced construct-irrelevant variance

^a Study has more than one focal and reference group, across grade. Sample sizes are averaged.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Writing Tools for English Language Arts

Students with disabilities use a wide variety of tools when writing and many of these tools are also used by students without disabilities. For example word processors, spell check, and grammar check are commonly used by individuals without disabilities as well as individuals who have disabilities that impact writing. Two other accessibility tools, speech to text and word prediction, are used by individuals with physical and learning disabilities but have become more widely used by individuals without disabilities in the form of both software (e.g., Dragon Naturally Speaking), built-in to hardware (e.g., the speech to text feature on iPhone, Siri), and auto correction for typing on smart phones. A review of the Smarter Balanced content specifications in ELA indicates some of these features may interfere with the measurement of ELA Claim 2 (i.e., “Students can produce effective writing for a range of purposes and audiences”) in lower grade levels (where “proper spelling and use of language conventions” are part of the construct being measured). For other ELA claims and the Mathematics claims, these tools and features do not appear to alter the construct but may have implications for scoring (e.g. rater bias). Since much of the research on these tools is integrated in the same studies (e.g., comparison of essays written with both grammar check and spell check compared to handwritten essays) we have combined all of these tools into a single section of this literature review. This section begins with a summary of research for each of the tools individually, followed by a collective discussion of the implications of these features on automated scoring engines.

Summary of Research on Writing Tools for ELA

Spell check.

There appears to be only one study (Hollenbeck, et al., 2002) that isolated the impact of spell check from word processing in an assessment context. This study found no significant differences between essays that were composed with a word processor without spell check and a word processor with spell check. It is worth noting that the lack of a significant difference may have been due to the small sample size (25 students per condition).

Grammar check.

There appears to be no studies specifically on the use of grammar check the context of assessment.

Word processor.

Several studies comparing handwritten and word processed essays were conducted in the between 1994 and 2001 and reviewed in Cahalan-Laitusis (2004). Three of the studies found that in general, human raters gave hand written essays higher scores than the same essays transcribed to computer. However, rating training was found to mitigate some of the difference in scores between hand written and word processed essays. In addition, two experimentally designed studies compared the impact of word processors on writing essays for students with disabilities. Findings from these studies were mixed with no significant differences in scores found in one study and results from the other study suggesting that students with disabilities performed worse when composing essays with a computer. Most of these studies were conducted more than a decade ago; these findings may not be generalizable to the present day given the vast changes in technology and computer usage. A recent study conducted by Lovett et al., (2010) examined the impact of extended time on both word

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

processed and handwritten essays composed by college students. Findings indicated that essay length was longer for word processed essays but not hand written essays.

Speech to text.

We reviewed two studies that examined the impact of speech to text technologies for generating essays in an assessment context. Both studies provided evidence that essays generated by speech to text were of higher quality than handwritten essays (for high school and college students with learning disabilities). However both studies had small sample sizes and did not compare the impact of speech to text technologies relative to the impact of other digital supports (e.g., word processing, grammar check, or spell check) which are more commonly used technologies for individuals with learning disabilities. The most recent study conducted by MacArthur and Cavalier (2004) also provided information on dictation accuracy rates which may be useful in determining the utility of this accommodation beyond students with prior experience using speech to text (a.k.a. dictation) software.

Word prediction.

While word prediction is becoming increasingly common as an accessibility feature, there are currently no published studies that examined the impact of word prediction in an assessment context.

Text to speech (for proofreading).

One study conducted by Raskind and Higgins (1995) examined the use of text to speech as a proofreading tool and indicated that students found more errors using text to speech than a human reader or no supports. This was a small study and conducted almost two decades ago so results may not be generalizable to current assessments.

Policy implications.

The limited amount of research on writing tools and the use of multiple writing tools simultaneously make it difficult to draw conclusions about individual features. In addition, the vast changes in technology usage over the last decade weaken the utility of earlier studies which were concerned about the confluence of word processing and writing skills. Consequently, it is advisable to focus policy decisions on the impact of writing tools on the construct being assessed by grade level rather than the results of the studies included in this literature review.

Implementation considerations.

Findings on how word processors affect writing test scores are limited. Research on test takers without disabilities indicates that highly experienced computer users tend to write better with a computer than by hand (Owston, Murphy, & Wideman, 1992; Russell & Haney, 1997; Russell & Plati, 2001). In all of these studies, the handwritten essays were transcribed to a typed format, so that raters only viewed typed essays. The most recent research on students without disabilities (Russell & Plati, 2001) was conducted with eighth- and tenth-grade students taking the language arts portion of the Massachusetts Comprehensive Assessment System (MCAS). This study indicated that the paper and pencil writing tests underestimate the performance of students who are accustomed to using a computer when writing (by 4 to 8 points on an 80-point scale). In addition several studies have evaluated the impact of typed and hand written essays on essay length and raters perceptions.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Annotated Bibliography for Writing Tools for English Language Arts

(61) Hollenbeck, K., Tindal, G., Harniss, M., & Almond, P. (2002). *The effect of spellcheckers on statewide writing test scores as an accommodation issue*. Eugene, OR: University of Oregon.

In this study, the authors examined the impact of spell check on essay quality. Seventh-grade students ($N = 50$) were administered the statewide writing assessment under one of two conditions: (a) word processor without spell checker and (b) word processor with spell checker. Results indicated that students in the word processor with spell checker group received significantly higher scores on the Oregon Statewide Writing Assessment composite score and on three traits: organization, sentence fluency, and conventions. No significant scoring differences were found for three other traits: ideas and content, voice, and word choice.

(62) Cahalan-Laitusis, C. (2004). *Accommodations on High Stakes Writing Tests for Students with Disabilities*. (ETS Research Rep. No. RR-04-13). Princeton, NJ: Educational Testing Service.

This report includes an overview of studies on writing accommodations conducted prior to 2003. Most of the studies examined difference between handwritten and word processed essays and the impact of rater bias on scores. Most of the studies on test takers without disabilities found that highly experienced computer users tend to write better with a computer than by hand (Owston, Murphy, & Wideman, 1992; Russell & Haney, 1997; Russell & Plati, 2001). In all of these studies, the handwritten essays were transcribed to a typed format, so that raters only viewed typed essays. The most recent research on students without disabilities (Russell & Plati, 2001) was conducted with eighth- and tenth-grade students taking the language arts portion of the Massachusetts Comprehensive Assessment System (MCAS). This study indicated that the paper-based writing tests underestimate the performance of students who are accustomed to using a computer when writing (by 4 to 8 points on an 80-point scale). In addition several studies have evaluated the impact of typed and hand written essays on essay length and raters perceptions. In addition, several of the reviewed studies examined the impact of rater bias. Researchers at ETS conducted a study comparing scoring of handwritten and word-processed essays (Powers, Fowles, Farnum, & Ramsey, 1994). Subjects in this study wrote one essay by hand and a second essay on a computer. All handwritten essays were transcribed into a computer, and all word-processed essays were transcribed by hand. Initial results confirmed earlier research indicating a rater bias against typed essays. This study was replicated after changes were made to the training of raters. These changes included using both handwritten and word-processed essays in training, emphasizing that handwritten and word-processed essays may make different impressions, acknowledging the influences of perceived length on essay scoring, and checking for differences in the scoring standards. After the revised training, the difference between scores on the handwritten and word-processed essays was smaller. The third study (Hollenbeck, Tindal, Stieber, & Harniss, 1999) compared the ratings of 80 essays that were originally handwritten as part of a middle school statewide writing assessment. Results indicated that scores on three of the traits (ideas and content, organization, and conventions) were significantly lower for the typed essays than for the hand-written essays. The results of these studies suggest that any mixed use of handwritten and word processed essays should be monitored during scoring. Studies of significance included in this synthesis were:

Hollenbeck, K., Tindal, G., Stieber, S., & Harniss, M. (1999). *Handwritten versus word-processed statewide compositions: Do judges rate them differently*. Eugene, OR: University of Oregon. Individuals with Disabilities Education Act of 1991, 20 U.S.C. § 1400 et seq. (1991).

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement, 31*, 220-233.

Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computers and via paper-and-pencil. *Educational Policy Analysis Archives, 5*(3). Retrieved September 25, 2002, from <http://epaa.asu.edu/epaa/v5n3.html>

Russell, M., & Plati, T. (2001). Effects of computer versus paper administration of a state-mandated writing assessment. *Teachers College Record*. Retrieved May 10, 2012, from <http://www.bc.edu/research/intasc/PDF/ComputerVsPaperStateWriting.pdf>

(63) Lovett, B. J., Lewandowski, L. J., Berger, C., & Gathje, R. A. (2010). Effects of response mode and time allotment on college students' writing. *Journal of College Reading and Learning, 40*(2), 64-79.

This study examined essay items for a college-based course examination; participants were screened using measures of writing, including brief sentence composition, motor speed, and essay writing skill: Woodcock-Johnson Tests of Achievement, Third Edition, Form A (WJ-III) (2001), writing fluency subtest; one-minute sample of writing to measure handwriting and typing speeds, and Test of Written Language, Third Edition (TOWL-3), Form B (1996). The results of the comparisons between accommodation conditions, and across the interactions, were complex across the performance tasks. Not surprisingly, students typed more words in the essay and speed tasks than they handwrote; however, there were no differences in quality between handwritten and word-processed responses. The extended time accommodation was associated with an increase in essay length, but only when produced with word processing, not handwriting. The length of essays was not related to their quality when handwritten, but longer word-processed essays scored higher.

(64) Higgins, E. L., & Raskind, M. H., (1995). Compensatory effectiveness of speech recognition on the written composition performance of postsecondary students with learning disabilities. *Learning Disability Quarterly, 18*, 407-418.

This study, conducted at the California State University – Northridge, examined the impact of speech to text on composition or editing of essays for college students with learning disabilities. In the study, 29 college students with learning disabilities were asked to write three essays, one for each of the following conditions: using a speech recognition system, dictating the essay to a human scribe, and without any assistance. Under the no assistance condition, students were allowed to handwrite or word-process their essay without using the spell check function. Essays were holistically scored on a scale of 1 to 6. Research findings indicated that speech recognition assists students with learning disabilities in compensating for their difficulties in written composition. Holistic scores for essays that were composed using speech recognition were significantly higher than the holistic scores achieved under the no assistance condition. The scribe condition was not significantly different from either the no assistance or speech recognition conditions. Limitations of this study are the small sample size and inconsistencies between formats (handwritten versus typed) that were scored.

(65) MacArthur, C. A., & Cavalier, A. (2004). Dictation and speech recognition technology as accommodations in large-scale assessments for students with learning disabilities. *Exceptional Children, 71*, 43-58.

This study examined the impact of speech recognition technology on the essays of high school students. The purpose of the study was to examine accuracy of the dictation software and not the

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

comparison of essay quality or differential boost between essays generated with and without speech to text technology. This was a relatively small study ($N = 30$) so results may not generalize to other student populations (particularly younger students). The results indicated that 68% of the students achieved 85% accuracy and 40% achieved 90% accuracy using dictation to a scribe or speech recognition software. However 3 students (10% of the sample) achieved below 80% accuracy. The authors reported that the results also demonstrated that both dictation conditions helped students with learning disabilities produce better essays (compared to handwritten essays).

(66) Raskind, M. H., & Higgins, E. L. (1995). Effects of speech synthesis on the proofreading efficiency of postsecondary students with learning disabilities. *Learning Disability Quarterly*, 18, 141-158.

This study focused on college student's ability to edit a previously written document under three different conditions (a) text-to-speech conversion technology, (b) a human reader, or (c) no assistance. In this study, 33 college students with learning disabilities were asked to write an essay either by handwriting or word processing without spell check, and then return for a second session to proofread and locate errors in their essays under one of the three conditions. The text-to-speech condition allowed the student to select text on a computer screen and hear the words spoken as they were simultaneously highlighted. Students were allowed to modify the rate of speech, volume, pitch, and background colors. No time constraints were placed under any of the three conditions. Results indicated that subjects found significantly more of the total errors in the text-to-speech condition (36% of errors were found) than in either the human reader condition (32%) or the no assistance condition (25%).

Table 6: Annotated Bibliography for Writing Tools

Study no.	Purpose Compare scores	Accommodations						Grade				Design			Writing tools		
		Word Processor	Spell Check	Grammar Check	Word Prediction	Text to Speech (for editing)	Speech to Text for dictation	Elementary	Middle	High	College	Experimental/quasi-Experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference	Major findings
61	✓	✓	✓					✓				✓			50	N/A	In Grade 7 students with disabilities who used word processor with spell checker received higher overall essay scores than students who used a word processor without spell checker. The impact on trait scores was mixed.
62	✓	✓						✓	✓			✓			N/A	N/A ^a	Students with experience using word processors write better essays on a computer than paper. In addition a rater bias exists between handwritten and typed essays which have implications for training of raters if paper forms are used.
63	✓	✓									✓	✓			140	N/A	College students with learning disabilities who received extended testing time generated longer essays when they used a word processor than when they hand wrote an essay. This has implications for scoring since longer essays received higher scores.
64	✓	✓				✓					✓	✓			29	N/A	College students with learning disabilities who used speech to text software for dictating an essay had higher scores than those who used a human scribe or a word processor with no dictation assistance.

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

Study no.	Purpose	Accommodations						Grade				Design			Writing tools		
		Word Processor	Spell Check	Grammar Check	Word Prediction	Text to Speech (for editing)	Speech to Text for dictation	Elementary	Middle	High	College	Experimental/quasi-Experimental	Operational data	Qualitative/descriptive	Sample size focal	Sample size reference	Major findings
65	✓	✓				✓					✓		✓	30	N/A	High school students with learning disabilities produced better essays with the use of speech to text software for dictating an essay than when they handwrote an essay. The accuracy rate of the speech to text software was variable between individuals.	
66	✓	✓			✓				✓		✓			33	N/A	College students with learning disabilities were able to identify more errors in a previously written document using of text-to-speech technologies compared to a human reader or no assistance.	

Literature Review of Testing Accommodations and Accessibility Tools for Students with Disabilities

References

- Altman, J. R., Lazarus, S. S., Quenemoen, R. F., Kearns, J. J., Quenemoen, M. M., Thurlow, M. L., & National Center on Educational Outcomes. (2010). Survey of States: Accomplishments and New Issues at the End of a New Decade of Change, 2009. *National Center On Educational Outcomes, University Of Minnesota*.
- Buzick, H., & Stone, E. (2011). *Recommendations for conducting differential item functioning (DIF) analyses for students with disabilities based on previous DIF studies*. ETS Research Report 11-34
- Ellington, A. J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education*, 34 (5), 433-463.
- Hess, K. (2011, September). Draft content specifications with content mapping for the summative assessment of the common core state standards for English Language Arts & Literacy in History/Social Studies, Science, and technical subjects. Retrieved from <http://www.k12.wa.us/SMARTER/ContentSpecs/ELA-LiteracyContentSpecifications.pdf>
- Scheuneman, J. D., Camara, W. J., Cascallar, A.C., Wendler, C., & Lawrence, I. (2002). Calculator access, use and type in relation to performance on the Sat I: reasoning test in mathematics. *Applied Measurement in Education*, 15(1), 95-112.
- Smarter Balanced Assessment Consortia (March, 2012) Content Specifications for the Summative assessment of the Common Core State Standards for Mathematics. Retrieved April 12, 2012 <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2011/12/MathContentSpecifications.pdf>