Podcasts on Mobile Devices as a Read-Aloud Testing Accommodation in Middle School Science Assessment

Don McMahon, Rachel Wright, David F. Cihak, Tara C. Moore & Richard Lamb

Journal of Science Education and Technology

ISSN 1059-0145

J Sci Educ Technol
DOI 10.1007/s10956-015-9591-3

Springer

J Sci Educ Technol
DOI 10.1007/s10956-015-9591-3

Podcasts on Mobile Devices as a Read-Aloud Testing Accommodation in Middle School Science Assessment

Don McMahon[1] · Rachel Wright[2] · David F. Cihak[3] · Tara C. Moore[3] · Richard Lamb[1]

Abstract The purpose of this study was to examine the effect of a digitized podcast to deliver read-aloud testing accommodations on mobile devices to students with disabilities and reading difficulties. The total sample for this study included 47 middle school students with reading difficulties. Of the 47 students, 16 were identified as students with disabilities who received special education services. Participants were randomly assigned to three experimental testing conditions, standard administration, teacher-controlled read-aloud in traditional group delivery format, and student-controlled read-aloud delivered as a podcast and accessed on a mobile device, and given sample end-of-year science assessments. Based on a factorial analysis of variances, with test conditions and student status as the fixed factors, both student groups demonstrated statistically significant gains based on their testing conditions. Results support the use of podcast delivery as a viable alternative to the traditional teacher-delivered read-aloud test accommodation. Conclusions are discussed in the context of universal design for learning testing accommodations for future research and practice.

Keywords Testing accommodations · Read-aloud · Mobile devices · Universal design for learning · Science education

✉ Don McMahon
macddm@gmail.com; don.mcmahon@wsu.edu

[1] Washington State University, Pullman, WA, USA

[2] University of Tennessee, Knoxville, TN, USA

[3] Department of Theory and Practice in Teacher Education, University of Tennessee, Knoxville, TN, USA

Introduction

With the introduction of both the No Child Left Behind Act (*NCLB*) in (2001) and the reauthorization of the Individuals with Disabilities Education Improvement Act (*IDEIA*) in (2004), students with disabilities must, to the greatest extent possible, be included in all large-scale, statewide testing programs, thus increasing schools' accountability for the academic achievement of all students. These policy initiatives have led to drastic increases in the number of students now required to participate in such assessments, largely a result of including students with disabilities who were previously excluded for accountability reasons. Therefore, accessibility is becoming an increasingly desired feature of accountability tests in the K-12 environment. In this new paradigm, students with various physical, cognitive, sensory, or linguistic barriers are given a less biased opportunity to demonstrate their knowledge, skills, and abilities on end of course measurements (Winter et al. 2006; Ketterlin-Geller et al. 2007; Beddow et al. 2009). As such, accessibility is one part of unified construct of validity, the degree to which the interpretation of a test score is justifiable for a particular purpose and supported by evidence and theory (AERA et al. 1999; Messick 1989; Laitusis et al. 2012). However, accessibility is not a static property, but rather represents the dynamic interactions among the items and features of a test and student characteristics, such as ability level and learning style preferences. Successful student responses, then, are either permitted or inhibited as a result of these ever-changing elements and interactions (Dolan and Rose 2000; Winter et al. 2006). Providing options for testing accommodations offers increased accessibility of assessments for students with disabilities.

Published online: 09 November 2015

Accommodations

Testing accommodations increase accessibility by removing construct-irrelevant variance caused by physical, cognitive, or sensory barriers preventing access to the material being assessed (Dolan et al. 2005) and, therefore, reducing error in the testing variance. Accommodations represent variations in the administration of assessments to moderate the effects of a student's disability that inhibit understanding or expression of content-specific knowledge. Common testing accommodations include the use of flexible scheduling and timing of the testing situation, variations in the test setting, revised test formats, variations in item presentation, and responses to attempt to provide educators with the best indication of students' content mastery possible (Haladyna and Downing 2004; Thurlow et al. 2006; Meloy et al. 2002).

Cormier et al. (2010) reported that one of the most frequently used types of testing accommodation is read-aloud, which typically involves a teacher providing a live reading of the test questions and answer choices to a student or group of students. This practice represents a variation in test administration, or presentation of items, and is a common application for students with learning disabilities in reading or for any students whose disabilities adversely affect the ability to independently read the test items and/or answer choices accurately (Johnstone et al. 2006). Intended to provide access to the content itself without interference from deficits in reading decoding and/or comprehension, read-aloud accommodations have been used for students with disabilities across elementary (Bolt and Thurlow 2007), middle (McKevitt and Elliott 2003), and high school grades (Elbaum 2007), as well as across content domains, including mathematics (Ketterlin-Geller et al. 2007), science (Meloy et al. 2002), and reading (Temple 2007) for students with reading difficulties and disabilities.

According to the differential boost framework (Fletcher et al. 2006; Fuchs 1999), an accommodation is justified when a student with a disability obtains a larger increase in score when compared to a student without a disability as a result of the accommodation, specifically. In a recent meta-analysis, Li (2014) analyzed 114 effect sizes from 23 studies conducted to examine read-aloud accommodations for students with and without disabilities and found that all students included in the review benefited from the read-aloud accommodations, regardless of disability status. However, for the students with disabilities, the accommodation effect size was significantly larger compared to the effect size for students without disabilities, which supports read-aloud as a successful accommodation in reducing barriers presented through disabilities on tests.

Although the use of read-aloud accommodations for certain content, such as reading assessments, has been a topic of debate among researchers, researchers have viewed read-aloud as a logical option when assessing mathematics or science concepts (Cline et al. 2008; Cook et al. 2008; Steinberg et al. 2008). For example, Meloy et al. (2002) examined the effects of administering assessments with and without read-aloud accommodations to middle school students with and without identified reading difficulties. Both student groups achieved significantly higher test scores with the read-aloud administration across assessments in both science and mathematics. According to previous read-aloud research (Tindal et al. 1998; Johnstone et al. 2006; Laitusis et al. 2012), the read-aloud accommodation supports students who have a reading deficit without giving them an undue advantage over those who do not receive the accommodation, which is the rationale underlying the use of all testing accommodations.

Although read-aloud accommodations have been shown to provide a more equal opportunity to demonstrate content knowledge for students who require such testing variations, discrepancies in read-aloud delivery, student processing differences, and the extensive amount of teacher time required have been identified as challenges that arise during implementation. First, there are several ways in which human error can occur when administering the read-aloud accommodations (Landau et al. 2003). Ketterlin-Geller et al. (2007) found that read-aloud practices varied considerably within localized areas. Some of the issues identified included teachers mispronouncing words, misreading words, verbalizing non-scripted comments, and/or making intonations that could inadvertently influence a student's attention or response. These variations in reading a test aloud raise concerns about both its reliability and validity unless standardized testing protocols are followed.

Second, read-aloud procedures impose a linear navigation path and an arbitrary set pace upon students as they work to understand and process test items. Students take the test in sequence rather than have the opportunity to skip to other test items, answer known questions first, or apply other test-taking strategies (e.g., reading possible answers before the question, skipping and coming back to difficult questions). Students may be reluctant or not permitted to ask a teacher to repeat selections of test items needed (Landau et al. 2003), as one student's request for an additional reading of an item is provided to the entire group. Thus, group read-aloud administration forces all students to respond to the content at the same teacher-controlled pace, rather than progressing through the assessment at each student's own self-selected pace.

Although a one-to-one (teacher-to-student) format offers a possible solution to meet each student's processing needs,

intensive teacher time is required, and teachers may quickly find this too time-consuming and unfeasible. One possible solution is the use of computer-based assessments. Computer-based assessments can provide text-to-speech features, increasing independent and self-paced access to the text; however, not all classrooms are equipped with the number of computers needed to implement computer-based testing (CBT) accommodations (Dawson et al. 2000; Farmer et al. 1992; Hebert and Murdock 1994; McCullough 1995; Strangman and Dalton 2005).

The combination of human error, linear progression of testing, and teacher time suggests a need for uniformly administered read-aloud testing accommodations. Technology-based solutions such as podcasts may be capable of providing the benefits of read-aloud accommodations to students with disabilities and reading difficulties while addressing some of these major issues challenging practical use. For example, Flowers et al. (2011) compared CBT and pencil-and-paper testing (PPT) read-aloud accommodations. Students in the PPT condition with an adult reader had higher mean scores in almost all academic content areas than those with the CBT read-aloud condition, although effect sizes ranged from small to moderate. However, students self-reported higher rates of performance when using the computer and an overall preference of the CBT condition were reported by both students and staff. It was suggested by the authors that additional research would be required in order to further evaluate the relation between testing modes and academic performance. In addition, Buzick and Stone (2014) analyzed 19 comparison studies with video using the read-aloud delivery mode. They concluded that, on average, computer read-aloud did not improve test scores on mathematics assessments taken by students with disabilities or students without disabilities. However, since a limited number of studies were identified using computer-based read-aloud assessments, Buzick and Stone noted that the results may not generalize to future uses of read-aloud delivered via text to speech. Both Li (2014) and Buzick and Stone (2014) suggested that read-aloud accommodations delivered via digital or text to speech will likely be the most common delivery mode of future assessments and recommended the need for more empirical studies.

Building on the universal design for learning (UDL) principles and guidelines (Rose and Meyer 2002; CAST 2011), we developed a podcast read-aloud test administration format, in which students heard their teacher's voice delivered through audio files enhanced with additional information, such as pictures and menus accessed through the use of a mobile device. As noted by other researchers, UDL is an important consideration in the development of accessible assessments (Thompson et al. 2002). Through the podcast format created, students were provided with the ability to control their own testing experience, by allowing self-regulation of the read-aloud pacing and order of items selected as needed.

The purpose of the study was to examine the effects of student-controlled podcasts on mobile devices as a read-aloud testing accommodation in comparison with the two more established testing conditions of standard administration (i.e., no accommodations) and traditional teacher-controlled read-aloud in a group delivery format. The sample included 47 middle school students who were a part of an intervention program for students with reading difficulties, 16 of whom had been identified as having a disability and regularly received read-aloud testing accommodations. Specifically, this study was guided by the following research questions: (1) Do statistically significant differences exist between student performance on science content tests when compared across the testing conditions: standard administration, teacher read-aloud, and podcast read-aloud? (2) Do significant differences exist across the testing conditions between the two groups of students, special education and general education?

Methods

Design

This study used a generalized Latin square comparative design to examine the differences within and between student groups across the three experimental testing conditions. The single factor of interest for the authors was student performance on a science content assessment. Within the Latin square design, main effects are assigned within each cell, and participants receive one dosing of the treatment and are compared. The authors acknowledge that the students were not isolated from the classroom environment; therefore, the selected design is not a traditional Latin square design. The ability to examine data collected within the natural classroom environment created more natural variance in the data structure and therefore was considered advantageous (Lamb et al. 2012; Lamb and Annetta 2013; Lamb et al. 2014). There are two main advantages to this type of research design. First, the researchers were allowed to examine each of the main effects independently and within the student groupings. Sensitization and carryover effects are mitigated via randomization of testing questions as described in the materials section. Time between administrations of each test was 3 days. Secondly, this design allowed for isolation of the treatment effects within fewer iterations and allowed all members of the sample to receive exposure to all three testing conditions. There is not a concern of ordering effect related to carry over or prior exposure due to the

complexity of the tasks and length of series in addition to time between administrations and randomization. Hogarth and Einhorn (1992) provide evidence that complexity or task, length of series, and time between administrations are sufficient to mitigate ordering effects. Testing for ordering effects via examination of significant reduction in error was also conducted. Results of the test for error reduction were not statistically significant, suggesting that there are no significant ordering effects (Carlsson et al. 2012). Table 1 provides an overview of the study design.

Analysis

A factorial analysis of variance (fANOVA) was employed to compare student performance across the three testing conditions and to measure the effect of status between the two groups of students (students with disabilities and students with reading difficulties). The use of fANOVA is indicated when using a Latin square design as a means to allow examination of relationships around multiple factors, while revealing any effects of interactions on individual factors that may have otherwise been hidden. Assumptions for fANOVA are the same as those for a traditional one-factor ANOVA. Repeated measures ANOVA was not used in this particular analysis as repeated measures ANOVA is used when investigations occur over three or more time points or multiple conditions for the same sample. The use of factorial ANOVA allows for examination of the interaction effects not otherwise visible during the use of repeated measures ANOVA. Lastly, given that repeated measures ANOVA is a special form, factorial ANOVA accounting for dependence between observations of the use of a factorial ANOVA assuming independence of observations provides more conservative examinations of effects and differences (Schörgendorfer et al. 2011).

Participants and Setting

The sample group for this study included 47 sixth-grade students with reading difficulties selected from four science classrooms in an urban middle school in the southeastern USA. The total sample was comprised of 16 students with disabilities who regularly received teacher read-aloud accommodations for testing, and 31 students identified as having reading difficulties. All of the students selected participated in a reading intervention program designed for students who were more than 2 years behind in benchmarked reading levels. Additionally, all students were placed in the first level of the reading intervention program, which indicated an independent text reading ability between first grade and third grade. Marked deficits in reading fluency prohibited the students from taking science and other content area tests independently. Of the total sample of participants ($N = 47$), 36 students were African American (76.6 %), eight were Caucasian (17.0 %), two were Hispanic (4.3 %), and one was Native American (2.1 %). Further, the sample was 57.4 % male and 42.6 % female. At the time of the study, the school was identified as a High Priority School because of a "needs improvement" score on performance from previous years on achieving adequate yearly progress (AYP), and 97.4 % of students were receiving free or reduced lunch prices, which was much higher than the state average of 56.0 %.

Collaboration took place between the researchers, a sixth-grade science teacher, and a special education teacher from the school to establish protocols for each of the three testing conditions. The Virginia read-aloud testing manual was adopted to serve as the primary reference for the instructions on reading complex science questions aloud (Virginia DOE 2006) to ensure each question was read the same way, whether it was recorded in the podcast or read by the teacher in the group administration condition.

Variables and Data Collection

Dependent Variable

The dependent variable for the study was the percent of questions answered correctly on a 30-question science test. Three versions of the test were created in order to mitigate any carry over effects. First, 90 multiple-choice questions were selected from a formative assessment series for sixth-grade science. Readability of test questions was assessed using the Fletch Kincaid (Kincaid et al. 1975) method, which indicated the questions were written at a 4.5 grade level. The 90 questions selected were then randomized and

Table 1 Experimental design

Group	Condition 1	Condition 2	Condition 3
Group 1	Standard administration	Podcast read-aloud	Teacher read-aloud
	Test version A	Test version B	Test version C
Group 2	Teacher read-aloud	Standard administration	Podcast read-aloud
	Test version B	Test version C	Test version A
Group 3	Podcast read-aloud	Teacher read-aloud	Standard administration
	Test version C	Test version A	Test version B

divided into three different tests of 30 items each. All three 30-item tests were independently reviewed by the research team members to ensure content topics were balanced (i.e., phases of the moon) and represented on all versions. The questions were designed to provide a measure of science knowledge similar to the statewide summative assessment format, which included multiple-choice questions with four answer choices. Students received a printed copy of the multiple-choice test under all three test conditions.

Independent Variables

Independent variables were the three testing conditions: (1) standard administration (no accommodation), (2) teacher-delivered read-aloud in group delivery format, and (3) podcast read-aloud on a mobile device. The established testing protocols were implemented to provide as much consistency as possible among the three testing conditions. Additionally, all testing was conducted in classroom environments familiar to the students under all three conditions, and the general classroom instructions for testing were employed (e.g., eyes on your own work, no talking, raise your hand if you have a question).

Standard Administration (no Accommodation Condition) This condition is the standard administration for classroom formative assessments with no accommodations. Students received a printed copy of the test and were instructed to complete it without assistance or use of any additional materials. Students were simply instructed to select the best answer to the question and to do their best work. Students were reminded not to talk during the test, and that the teacher would not be able to help them with difficult words.

Teacher Read-Aloud Accommodation The teachers providing teacher read-aloud were instructed to treat this situation as if it were a chapter or unit test in their classroom. Additional student instructions were provided to ask the teacher to repeat a question, if needed, by raising hands. In addition to the established protocols, the teacher read the test according to the school district's read-aloud procedures, as follows: read each question and answer choices once, wait approximately 10–15 s, and then ask whether everyone is ready to hear the next question. If no requests for repeated reading of the question were received, the teacher moved on to read the next question sequentially through the end of the test.

Podcast Read-Aloud Accommodation An enhanced podcast, which incorporated chapter markers making each question its own "chapter," was added to the recording of the test to create a final product that was playable on a variety of mobile devices. This allowed the student to play, listen, pause, repeat, or advance the question as needed. All students were provided with the same recording of a teacher from their locality employed by the school district reading test questions. This reader was trained by the district in how to read tests aloud according to the state's testing accommodations manual. Using the mobile devices to access the podcast gave students the ability to self-operate and self-manage the delivery of the read-aloud accommodation as well as control the order and pace of test questions read. The podcasts allowed the students to best meet their individual preferences and cognitive processing needs. The podcasts were played aloud on iPod touch devices. During the podcast read-aloud condition, students were provided an iPod touch with the podcast on it and a set of headphones. Further instructions were provided under the podcast condition instructions for using the mobile device. These included how to access the read-aloud test, navigate through test items, and general information such as "the device may only be used to listen to the podcast" and "raise your hand for help if you have problems using the device."

Materials

The podcast version of the test was produced in software called Garageband (Apple, 2009). Using Garageband, the questions were recorded and edited to create an enhanced podcast-audio version of the test questions. An enhanced podcast is a digital recording of a human voice, which can be created using a variety of software tools, and allows the user to insert a variety of additional content including pictures and video. The files are produced so that each question is clearly labeled and selectable by the student. All three test versions were read by the same reader and were read without emphasis and without the reader having direct knowledge of the correct answers to the questions. The audio of each question was digitally set as a chapter on the audio file of the enhanced podcast. Then, each "chapter" was labeled with a graphic label according to the question that was read aloud. See Fig. 1. This allowed students to use a menu to select the question to be read. Using the menu feature built into the edited file, students could navigate or skip to various question items on the test.

Results

The results of the fANOVA omnibus test indicate that there are statistically significant main effects related to testing conditions $F(1,45) = 6.23, p = .016, \eta^2 = .13$ (large). The large effect size suggests that testing conditions account for a significant amount of the variance between groups in the

Fig. 1 Podcast read-aloud test on an iPod Touch

study. η^2 is the analogue of R^2 in multiple regressions. In addition, interactions between special education student status and test type were found to be statistically significant, $F(1,45) = 8.72$, $p = .001$, $\eta^2 = .273$. This result suggests that the type of testing condition (teacher read-aloud, podcast, or standard administration) affected the performance of special education students.

Eta squared for omnibus test is .13, indicating that 13 % of the differences in special education student performance is due to test type. Examination of partial eta squared for the interaction of special education and test type suggests that 27.3 % of the differences within the special education group are accounted for by differences among groups holding test type constant.

The Tukey's post hoc method of comparisons was used to determine differences among testing conditions between groups, due to statistically significant difference between the independent testing condition and the podcast testing condition ($p < .001$). The results provided by the 95 % CI for the difference between the means of the testing type indicate that the podcast testing conditions result in difference between 4.7 points and 19.11 points. Additionally, the post hoc test suggests that the teacher read-aloud condition resulted in statistically significant gains when compared to the standard testing condition resulting in a difference between 1.6 points and 13.6 points. It should also be noted that the difference between the teacher read-aloud and podcast conditions was not statistically significant.

Results by Testing Condition

The mean for the standard administration (no accommodation) condition was 46.30 % correct (SD = 22.89), teacher-controlled read-aloud condition was 54.46 % correct (SD = 20.07), and the podcast condition was 57.46 % correct (SD = 19.87) for the entire sample ($N = 47$). The results also appear to have relatively large standard deviations. See Table 2 for results.

Podcast Versus Standard Administration (no Accommodation)

The overall mean difference in scores achieved between the podcast and standard administration (no accommodation) condition was +11.16 % points (SD = 18.75). The data support our research hypothesis that podcast delivery of read-aloud testing accommodations increases student performance scores, $t(1, 46) = 4.08$, $p < .001$. Cohen's $\hat{d}$, a measure of the effect size of the treatment, indicated that the podcast read-aloud produced a positive effect of medium size on the mean scores, $d = .59$ (Cohen 1988).

Teacher Versus Standard Administration (no Accommodation)

The overall mean difference in scores between the teacher-controlled read-aloud condition and standard administration was +8.16 % points (SD = 15.65). The data support the hypothesis that teacher read-aloud testing accommodations increased achievement scores, $t(1, 46) = 3.57$, $p = .001$. Cohen's $\hat{d}$ indicated that the positive effect of teacher read-aloud produced a medium effect size on the student achievement scores ($\hat{d} = .52$).

Podcast Versus Teacher-Controlled

The mean difference in the podcast read-aloud compared to teacher-controlled read-aloud was +3.00 % points

Table 2 Pairwise comparisons for each of the three testing conditions: standard administration, teacher read-aloud, podcast read-aloud for all students $n = 47$

Comparisons	Mean difference	SD	*t*	*df*	Sig.
Podcast–independent	11.16	18.75	4.08	46	.000
Teacher–independent	8.16	15.65	3.57	46	.001
Podcast–teacher	3.00	16.48	1.24	46	.218

(SD = 16.48), which was not significantly different, $t(1, 46) = 1.25$, $p = .218$.

Pairwise Comparisons by Student Group

The overall mean scores for the three testing conditions were compared according to assigned student group: students with disabilities ($n = 16$) and students without disabilities but identified as having reading difficulties ($n = 31$). Using pairwise comparisons, the results of each testing condition across the two student groups are reported below. See Table 3. Levene's tests for equal variances were conducted on each comparison revealing equal variance could be assumed for all of the following comparisons. As significance was found, the statistical test exhibited sufficient power to detect differences based on the current sample size.

Under the standard administration (no accommodation) testing condition, the students with reading difficulties scored significantly higher ($M = 51.06$ % correct, SD = 24.56) than students with disabilities ($M = 37.08$ % correct, SD = 16.23), $t(1, 45) = 2.05$, $p = .046$. Cohen's $\hat{d}$, a measure of the effect size, indicated that the student status with no accommodations accounted for a medium effect size on the mean scores ($\hat{d} = .61$).

Under the teacher-controlled group, read-aloud accommodation condition, students with reading difficulties scored significantly higher ($M = 60.32$ % correct, SD = 20.19) than students with disabilities ($M = 43.12$ % correct, SD = 14.57), $t(1, 45) = 1.54$, $p = .004$. Cohen's $\hat{d}$, a, indicated that the student status with teacher read-aloud accounted for a small effect size on the mean scores ($\hat{d} = .45$).

Under the podcast read-aloud testing accommodation condition, there was no significant difference in scores for students with reading difficulties ($M = 60.64$ % correct, SD = 21.11) and students with disabilities, (51.31 % correct, SD = 16.07), $t(1, 45) = 1.54$, $p > .05$ ($p = .128$). Results are summarized in Fig. 2.

Discussion

These results suggest that podcast read-aloud condition was a viable alternative to teacher read-aloud. These results support the hypothesis that podcast read-aloud would produce an increase in scores compared to the no accommodation condition does not alter student achievement significantly compared to teacher read-aloud condition. This study demonstrates a method for making assessments in science more inclusive for students who struggle with reading fluency. This study was consistent with Meloy et al. 2002, and the study is also consistent with recently completed meta-analyses by Li (2014) and Buzick and Stone (2014), which concluded that both student groups improved using the podcast read-aloud accommodation. Since the read-aloud accommodation yielded higher test scores for both groups of students, this indicates that the read-aloud served as an advantage to individuals with whom it was used. This result is contrary to what researchers and policymakers would intend for an "appropriate" accommodation. Many researchers and policymakers posit that an accommodation is justified if it positively affects the performance of students with disabilities and is neutral for students without disabilities. One seemingly obvious explanation for the higher scores under the read-aloud condition in the present study is the content assessed reading ability in addition to the science content skills and knowledge the test was designed to measure. Using multiple-choice items unavoidably places some reading demands on the student in order to measure science content knowledge and skills, despite attempts to minimize the possible influence of reading skills on the science achievement test. Another explanation for the higher test scores of both groups includes the possibility that the podcast read-aloud delivery system "(a) assisted students in maintaining their attention and reduced the chance of or need for skipping items, which would lower their scores; (b) permitted additional total test time for students to consider their responses; (c) allowed those who normally

Table 3 Pairwise comparisons of the means for general education students and special education students in each of the three testing conditions

Testing condition	Students status general ed. w/reading difficulties (GEN ED) or special ed. (SPED)	Mean	Mean difference	SE	*t*	*df*	Sig.
Independent	Gen ED	51.06	13.981*	6.814	2.05	45	.046
	SPED	37.08					
Teacher read-aloud	Gen ED	60.32	17.198*	5.698	3.02	45	.004
	SPED	43.12					
Podcast read-aloud	Gen ED	60.64	9.332	6.025	1.58	45	.128
	SPED	51.31					

* $p > .05$

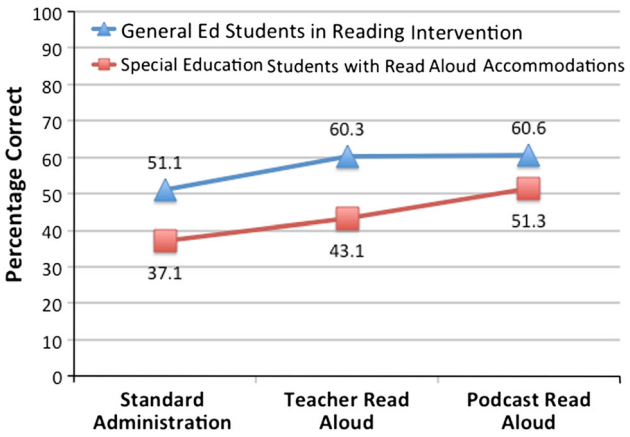

Fig. 2 Average achievement across testing conditions based on student status

work at a slower pace to not be discouraged by the quicker pace of many of their peers; or (d) unintentionally cued students to answers based on the reader's expressive style" (Meloy et al. 2002, p. 254).

UDL

The results of this study show significantly improved performance when the UDL principles were applied to the delivery of assessments. The results of this study suggest that the content knowledge of the students in special education with read-aloud testing accommodations was better measured when provided multiple representations of the assessment question. Only presenting the science questions in written form leads to significantly lower achievement for these students when compared to either teacher read-aloud or podcast read-aloud. Additionally, this implementation created benefits for students with reading difficulties who were in a remedial reading program even though they did not regularly receive read-aloud testing accommodations. The UDL approach of multiple representations allowed both groups of students to access the test content.

The podcast read-aloud solution examined in this study embodies many of the principles of UDL. The UDL principle (CAST 2009) provides multiple means of representation, and its guideline "provide options for perception" is demonstrated by the alternative to visual information provided by both podcast read-aloud and teacher read-aloud, but many other UDL guidelines are only demonstrated by the podcast delivery. In terms of the second UDL guideline "provide multiple means of action expression," the podcast delivery of read-aloud accommodations demonstrates several of these guidelines and checkpoints. For example, the sixth UDL guideline *options for executive functions* is demonstrated in the checkpoint *options that support planning and strategy development* by the students' ability to apply advanced test-taking strategies such as completing easy questions first. In traditional teacher read-aloud, students are limited to answering the questions in the order in which the teacher reads them. Finally, the third UDL principle of multiple means of engagement is demonstrated through several checkpoints including *options that increase autonomy* (they are less dependent on teachers) and *options that vary levels of challenge and support* (students choose what information they need read to them).

Podcast Student Advantages

Jablonski et al. (2008) found that students with disabilities achieve higher scores on assessments when they can set their own pace with a read-aloud accommodation. The podcast read-aloud allowed for more time for students to process questions at their own pace without verbal interruptions. Students were able to manage their own time better during tests since they were able to choose which questions to hear, how often they wanted to hear each question, and the order in which they heard specific questions. Hollenbeck et al. (2000) noted that student-controlled pacing, control over the speed in which a task is presented, increased persistence in difficult tasks, and that material read-aloud promoted higher levels of active engagement. Podcast read-aloud was tailored to the individual needs of the student. The majority of classroom instruction was lecture while the majority of testing was based on independent silent reading of the examination (standard administration of a test). The dichotomy between these two conditions could impact student achievement.

One shortcoming of the current teacher read-aloud accommodation relates to continuity between instruction and assessment (Landau et al. 2003). In order to receive a read-aloud accommodation, students are often removed from the setting where instruction occurred to have the test read-aloud. In the interest of keeping students in their least restrictive environment (LRE), (IDEA 2004) allowing students to take their tests in their general education classroom is a means of increasing the amount of time they are in inclusive settings. Students with disabilities were able to spend more time in general education classes including taking their tests with their peers in their instructional setting rather than being pulled out to have a test read to them. The student with read-aloud testing accommodations was no longer dependent upon a teacher reading the question to them. Using a mobile device, each student could listen to the parts of the test they need read-aloud, play a question according to their individual needs, or replay a question just like rereading a question. Learners using podcast read-aloud accommodations for testing are free to process the material in the questions at different

rates according to their unique educational strengths. One teacher reading to them all at the same time does not allow for individualized pacing options.

Teacher Advantages

The intervention provided teachers with increased time for instructional tasks since they did not have to read tests to students. Additionally, teachers did not have to read the same tests on multiple occasions because of student absences. The read-aloud accommodation delivered through podcasts returned time to teachers for instructional planning and tasks. Testing accommodations were more consistent and were not dependent on limited staffing.

In addition, it is not an effective use of teacher time to have every sixth-grade special education teacher read the test in light of the fact that the same test can be produced and shared securely with teachers to use when they need it for a student who has read-aloud accommodations. Teachers can easily convert any test into a format that can display images simultaneously with audio by teaching teachers to create math and science podcast read-aloud exams. Although teacher time is required to create a podcast read-aloud test, the amount of time saved by not having to read the test is likely to be greater since teachers can share audio versions of the test and will not have reread tests aloud a second time whenever a student is absent.

Teachers anecdotally reported that student behavior during testing situations was more positive using the podcast versus their behavior during standard and teacher read-aloud testing. Students were able to manage their own time better during tests since they were able to choose which questions to hear; consequently, less student waiting occurred. This process was faster than having to listen to the entire test being read by the teacher reading at the pace of the slowest test-taking student on each question.

Limitations

Student mean test scores for each of the testing conditions were relatively low because of the type and difficulty of the sample summative assessment questions. The difficulty of the questions both in readability and in content for this group of students is a limitation on analyzing the effectiveness of this intervention. The subject focus of science does not allow for this study to generalize the results to other core subjects. Also, the small sample size of the study limits the ability to generalize the results. While the use of questions based on high stakes tests added relevancy to the study, the results appear limited by the school's historically poor performance on annual yearly progress AYP measures. The decision to use question formats based on the high stakes test may have added difficulty to the trials in all of the conditions. Based on score averages, it appears that the questions were too difficult. Future studies should use a variety of question formats, subject areas, and designs in order to further examine read-aloud testing delivery options.

Future Research

Continued research is needed in order to better understand the accessibility and validity of assessments used with students with disabilities. As the number of students with disabilities in general education classrooms increases, the challenges of providing testing accommodations and modifications will also increase (VanGetson and Thurlow 2007). While beyond the scope of this study, the authors highly suspect that a full account of how many students with disabilities are consistently receiving their testing accommodations year round in all subjects would be very enlightening. We suspect there are simply not enough personnel available to provide the consistent read-aloud testing required by IDEA (2004). If our suspicion is empirically proven in future research, technology-based methods of read-aloud accommodations may prove to be one of the few practical solutions to meeting this need.

The researchers plan to apply this delivery of testing accommodations to the subjects of math and social studies. Building on the principles of UDL, especially "provide options for perception" (CAST 2012), the researchers hope to explore what factors increase students' ability to demonstrate their content knowledge. Additionally, future research should also address student preferences and achievement between a human recording like our research design of podcast read-aloud using enhanced podcasts, other delivery methods including text-to-speech options such as screen readers or the accessibility features available on some mobile devices.

Future research on this project can refine protocols on producing podcast read-aloud versions of assessments and conduct research on their impact on student achievement, teacher time, and fidelity of implementation.

Conclusion

This project represents a very feasible option to return teacher time spent reading tests to other instructional activities. Rather than having dozens of teachers reading the same test multiple times, a more effective use of resources may be to produce read-aloud podcast recordings of common district tests. Podcast read-aloud could allow students who have below grade level reading fluency to demonstrate content knowledge in the curriculum.

References

American Educational Research Association (AERA), American Psychological Association (APA), the National Council on Measurement in Education (NCME) (1999) Standards for educational and psychological testing. AERA, Washington

Beddow PA, Elliott SN, Kettler RJ (2009) Test accessibility and modification inventory (TAMI) accessibility rating matrix. Vanderbilt University, Nashville

Bolt SE, Thurlow ML (2007) Item-level effects of the read-aloud accommodation for students with reading disabilities. Assess Eff Interv 33(1):15–28

Buzick H, Stone E (2014) A meta-analysis of research on the read aloud accommodation. Educ Meas Issues Pract 33(3):17–30

Carlsson F, Mørkbak MR, Olsen SB (2012) The first time is the hardest: a test of ordering effects in choice experiments. J Choice Model 5(2):19–37

Center for Applied Special Technology (2011) Universal design for learning (UDL) guidelines: full-text representation (Version 2.0). http://www.udlcenter.org/aboutudl/udlguidelines

Cline F, Cook LL, Stone E (2008) An examination of differential item functioning on grade 5 math and science assessments for students with disabilities. Paper presented at the 2008 Annual Meeting of the American Educational Research Association, New York, NY

Cohen J (1988) Statistical power analysis for the behavioral sciences. Erlbaum, Hillsdale

Cook LL, Eignor DR, Steinberg Y, Sawaki Y (2008) Using factor analysis to compare the internal structure of a state standards-based math assessment. Paper presented at the 2008 Annual Meeting of the American Educational Research Association, New York, NY

Cormier DC, Altman J, Shyyan V, Thurlow ML (2010). A summary of the research on the effects of test accommodations: 2007–2008. (Technical Report 56). National Center on Educational Outcomes, Minneapolis, MN

Dawson L, Venn M, Gunter PL (2000) The effects of teacher versus computer reading models. Behav Disord 25(2):105–113

Dolan RP, Rose DH (2000) Accurate assessment through universal design for learning. J Spec Educ Technol 15(4):47–51

Dolan RP, Hall TE, Banerjee M, Chun E, Strangman N (2005) Applying principles of universal design to test delivery: the effect of computer-based read aloud on test performance of high school students with learning disabilities. J Technol Learn Assess 3(7):3–31

Elbaum B (2007) Effects of an oral testing accommodation on the mathematics performance of secondary students with and without learning disabilities. J Spec Educ 40(4):218–229

Farmer ME, Klein R, Bryson SE (1992) Computer- assisted reading: effects of whole-word feedback on fluency and comprehension in readers with severe disabilities. Remedial Spec Educ 13(2):50–60

Fletcher JM, Francis DJ, Boudousquie A, Copeland K, Young V, Kalinowski S, Vaughn S (2006) Effects of accommodations on high stakes testing for students with reading disabilities. Except Child 72(2):136–150

Flowers C, Kim D, Lewis P, Davis CA (2011) A comparison of computer-based testing and pencil-and-paper testing for students with a read-aloud accommodation. J Spec Educ Technol 26(1):1–12

Fuchs LS (1999) Curriculum-based measurement: updates on its application in standards-based assessment systems. Council for Exceptional Children, Charlotte

Haladyna TM, Downing SM (2004) Construct-irrelevant variance in high-stakes testing. Educ Meas Issues Pract 23(1):17–27

Hebert BM, Murdock JY (1994) Comparing three computer-aided instruction output modes to teach vocabulary words to students with learning disabilities. Learn Disabil Res Pract 9(3):136–141

Hogarth RM, Einhorn HJ (1992) Order effects in belief updating: the belief-adjustment model. Cogn Psychol 24(1):1–55

Hollenbeck K, Rozek-Tedesco MA, Tindal G, Glasgow A (2000) An exploratory study of student-paced versus teacher-paced accommodations for large-scale math tests. J Spec Educ Technol 15(2):29–38

Individuals with Disabilities Education Improvement Act, 20 U.S.C. §1400 (2004)

Jablonski B, Potts E, Wiley A (2008) Providing access to assessment: how should IEP teams make decisions about accommodations? http://www.cec.sped.org/AM/Template.cfm?Section=CEC_Today1&TEMPLATE=/CM/ContentDisplay.cfm&CONTENTID=11472

Johnstone CJ, Altman J, Thurlow ML, Thompson SJ (2006) A summary of research on the effects of test accommodations: 2002 through 2004. (Technical Report No. 45). National Center on Educational Outcomes, Minneapolis, MN

Ketterlin-Geller LR, Alonzo J, Braun-Monegan J, Tindal G (2007a) Recommendations for accommodations: implications of (in)-consistency. Remedial Spec Educ 28(4):194–206

Ketterlin-Geller LR, Yovanoff P, Tindal G (2007b) Developing a new paradigm for conducting research on accommodations in mathematics testing. Except Child 73(3):331–347

Kincaid JP, Fishburne RP, Rogers RL, Chissom BS (1975) Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel. (Research Brank Report 8-75). Chief of Naval Technical Training, Naval Air StationMemphis, Millington, Tenn

Laitusis C, Buzick H, Stone E, Hansen E, Hakkinen M (2012) Literature review of testing accommodations and accessibility tools for students with disabilities. Educ Test Serv, Princeton

Lamb RL, Annetta L (2013) The use of online modules and the effect on student outcomes in a high school chemistry class. J Sci Educ Technol 22(5):603–613

Lamb RL, Annetta L, Meldrum J, Vallett D (2012) Measuring science interest: Rasch validation of the science interest survey. Int J Sci Math Educ 10(3):643–668

Lamb RL, Vallett D, Annetta L (2014) Development of a short form measure of science and technology self-efficacy using Rasch analysis. J Sci Educ Technol 23(5):641–657

Landau S, Russell M, Gourgey K, Erin J, Cowan J (2003) Use of the talking tactile tablet in mathematics testing. J Vis Impair Blind 97(2):85–96

Li H (2014) The effects of read-aloud accommodations for students with and without disabilities: a meta-analysis. Educ Meas Issues Pract 33(3):3–16

McCullough CS (1995) Using computer technology to monitor student progress and remediate reading problems. Sch Psychol Rev 24(3):426–439

McKevitt BC, Elliott SN (2003) Effects and perceived consequences of using read-aloud and teacher-recommended testing accommodations on a reading achievement test. Sch Psychol Rev 32(4):583–600

Meloy L, Deville C, Frisbie DA (2002) The effect of a read aloud accommodation on test scores of students with and without a learning disability in reading. Remedial Spec Educ 23(4):248–255

Messick S (1989) Validity. In: Linn RL (ed) Educational measurement. Macmillan, New York

National Center for Educational Statistics (2006) The nation's report card: national and state reports in science now available. http://nationsreportcard.gov/science_2005/

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat.1425 (2002)

Rose DH, Meyer A (2002) Teaching every student in the digital age: universal design for learning. Association for Supervision and Curriculum Development, Alexandria

Schörgendorfer A, Madden LV, Bathke AC (2011) Choosing appropriate covariance matrices in a nonparametric analysis of factorials in block designs. J Appl Stat 38(4):833–850

Steinberg J, Cline F, Sawaki Y (2008) Examining the factor structure of a state standards-based assessment of science for students with disabilities. Paper presented at the 2008 Annual Meeting of the American Educational Research Association, New York, NY

Strangman N, Dalton B (2005) Using technology to support struggling readers: a review of the research. In: Edyburn D, Higgins K, Boone R (eds) Handbook of special education technology research and practice. Knowledge by Design Inc, Whitefish Bay, pp 545–569

Temple C (2007) The effects of interactive read-alouds on the reading achievement of middle grade reading students in a core remedial program. (Doctoral dissertation). Retrieved from Dissertation Abstracts International. (Accession Order No. 3287381)

Thompson SJ, Johnstone CJ, Thurlow ML (2002) Universal design applied to large-scale assessments (Synthesis Report 44). National Center on Education Outcomes, Minneapolis, MN. http://www.cehd.umn.edu/NCEO/onlinepubs/synthesis44.html

Thurlow ML, Thompson SJ, Lazarus SS (2006) Considerations for the administration of tests to special needs students: accommodations, modifications, and more. In: Downing SM, Haladyna TM (eds) Handbook of test development. Lawrence Erlbaum, Mahwah, pp 653–673

Tindal G, Heath B, Hollenbeck K, Almond P, Harniss M (1998) Accommodating students with disabilities on large-scale tests: an experimental study. Except Child 64(4):439–450

VanGetson GR, Thurlow ML (2007) Nearing the target in disaggregated subgroup reporting to the public on 2004–2005 assessment results. (Technical Report 46). National Center on Educational Outcomes, Minneapolis, MN

Virginia Department of Education (2006) Guidelines for administering the read-aloud accommodation for standards of learning assessments. http://www.doe.virginia.gov/testing/participation/guidelines_read_aloud.pdf

Winter PC, Kopriva RJ, Chen C-S, Emick JE (2006) Exploring individual and item factors that affect assessment validity for diverse learners: results from a large-scale cognitive lab. Learn Individ Differ 16(4):267–276