Article

Hammill Institute on Disabilities

Journal of Learning Disabilities
1–12

DOI: 10.1177/0022219417714774
journaloflearningdisabilities.sagepub.com
SAGE

Enhancing Accessibility for Students With Decoding Difficulties on Large-Scale Reading Assessments

Allison Gruner Gandhi, EdD[1], Burhan Ogut, PhD[2], Laura Stein, MA[1], Robin Bzura, BS[1], and Louis Danielson, PhD[2]

Abstract

This study reports findings from studies examining potential read-aloud accommodations on standardized reading comprehension assessments for students with decoding difficulties. Three types of accommodations were evaluated: question stems and answer options read aloud; question stems, answer options, and proper nouns read aloud; and full read-aloud. Drawing from a sample of 207 fourth-grade students with and without decoding difficulties, we used 3-level hierarchical linear modeling to assess whether there were significant differences between students with and without decoding difficulties in the effect of each accommodation relative to no accommodation. Analyses showed that, for the students with decoding difficulties, all 3 read-aloud accommodations had significant effects when compared with no accommodation at all, and these effects were significantly larger than for average readers (effect sizes: 0.49, 0.55, and 0.50, respectively). When non–individualized education program students were excluded from the group of students with decoding difficulties, the effect sizes became notably larger.

Keywords

accommodations, assessment, decoding, reading

Recent policy developments in large-scale assessment have important implications for ensuring that students with decoding difficulties are provided opportunities to successfully demonstrate their reading skills and knowledge. First, the widespread implementation of new assessments aligned to the Common Core State Standards (www.corestandards.org) brings an opportunity for greater consistency in states' policies on allowable test accommodations for students with disabilities, including those with decoding challenges. Second, the recent elimination of the "2% rule," which had permitted states to assess some students with disabilities against modified achievement standards, makes the need for valid accommodations even greater.

According to the International Dyslexia Association (2008), approximately half of students who qualify for special education are classified as having a learning disability (i.e., 6%–7% of the school population); about 85% of these students have a primary disability related to reading and language processing. Many more students, potentially as many as 15% to 20% of the population, have symptoms of dyslexia, depending on how it is defined (International Dyslexia Association, 2008; Shaywitz, 2003). Students with word-reading disabilities demonstrate difficulties that hinder accurate, fluent word reading and can result in poor spelling and poor writing. If students cannot read text accurately and fluently, then text comprehension is significantly compromised.

For students who have severe challenges decoding text, their word-reading accuracy interferes with measuring their comprehension and understanding of text. The challenge is to design test accommodations that do not lower the construct validity of the comprehension portion of the test or yield scores that misrepresent the comprehension construct, often referred to as *construct irrelevant variance* (Sireci, 2004). The goal of any accommodation is to minimize the impact of the modification of the assessment procedure without changing the measurement of the construct of interest, possibly leading to mistaken policy decisions or expectations (Fuchs, Fuchs, & Capizzi, 2005). Many researchers in the area of assessment have remarked on the need for further study of accommodations, including accommodations that involve reading aloud to students with reading disabilities on large-scale reading assessments (e.g., Lazarus et al.,

[1]American Institutes for Research, Waltham, MA, USA
[2]American Institutes for Research, Washington, DC, USA

Corresponding Author:
Allison Gandhi, American Institutes for Research, 201 Jones Rd, Waltham, MA 02451, USA.
Email: agandhi@air.org

2006; Thurlow et al., 2009). This study describes the impact of three accommodation types on reading comprehension test performance for students with decoding difficulties.

Policy Context

Beginning in the 2014–2015 school year, many states adopted new assessments aligned to a more rigorous set of college- and career-ready standards, including the Common Core State Standards. Most states adopted the assessments developed by one of the two major state testing consortia—the Partnership for Assessment for Readiness for College and Career (PARCC) or the Smarter Balanced Assessment Consortium (SBAC). Other states adopted a more rigorous version of their previous state assessments. These new assessments presented opportunities for states to reconsider their policies around testing accommodations. Importantly, the adoption of a common set of assessments across many states may facilitate consistency in accommodation policy, something that had previously been lacking.

PARCC and SBAC developed guidance on their testing accommodations policy for students with reading-related disabilities. PARCC (2015) specifically mentions accommodations to be provided for passages and literacy assessments, and it defines the type of students who would need accommodations as the ones "with print-related disabilities who would otherwise be unable to participate in the assessment because their disability *severely limits or prevents* their ability to access printed text by decoding" (p. 33). SBAC (2015) permits read-aloud of the full passage for students whose need is documented in an individualized education program (IEP) or 504 plan (p. 20). The full read-aloud accommodation in particular has been controversial, partly because it may change the focus of the assessment from reading comprehension to listening comprehension.

Perhaps as a result of the general guidance provided by these policies, there remains widespread diversity in practices surrounding reading assessment for students with reading disabilities, allowing states to identify their own policies surrounding test accommodations. Furthermore, guidelines on how to identify students who might benefit from such accommodations are vague: For the most part, accommodation decisions are made on a case-by-case basis, and it can be challenging for states to ensure that the use of the full read-aloud accommodation is constrained to only a small number of students.

Additionally, there is considerable uncertainty about the best way to provide test accommodations for these students (Laitusis, 2008). In some cases, students take their reading tests in an alternative setting, such as a small group, or receive extended time for testing. In other cases, portions of the test are read aloud, ranging from only the directions to the questions and/or answer choices to the full reading passages (e.g., Lazarus et al., 2006; Thurlow et al., 2009). Often, these accommodations are bundled for individual students.

Another relevant policy development is the recent elimination of what had previously been known as the *alternate assessment based on modified achievement standards* (AA-MAS). The AA-MAS was an option for students with persistent academic difficulties, including those with word-reading disabilities, and it was typically known as the "2% rule" because schools were permitted to count AA-MAS results toward their accountability determinations for up to 2% of their student population. Under the AA-MAS, students could take assessments with modified features, such as fewer response options, fewer items, fewer passages, simplified language, or visual emphasis on key words (Hodgson, Lazarus, & Thurlow, 2010). Growing concerns about the AA-MAS, which had been permitted since 2007, led the U.S. Department of Education (2013) to issue a statement that the use of the AA-MAS was preventing students with disabilities from reaching their true potential, and this rule was eliminated beginning in the 2014–2015 school year.

Without the AA-MAS, the use of valid test accommodations takes on even greater importance (Jamgochian & Ketterlin-Geller, 2015). According to an analysis conducted by Lazarus, Cormier, and Thurlow (2011), states with more restrictive accommodations policies were more likely to develop an AA-MAS than states that allowed a broader range of accommodations. Students in states that had previously used the AA-MAS may have less access to valid accommodations that could benefit them. More empirical evidence demonstrating the validity and benefits of testing accommodations for students with reading-related disabilities is critically needed.

Previous Research

Although numerous studies have been conducted on testing accommodations, variation in content focus, scope of accommodations reviewed, and results have made it difficult for experts to reach consensus (Lai & Berkeley, 2012). In general, research on accommodations has focused on testing the interaction hypothesis, which proposes that test accommodations should lead to improved test scores for students who need the accommodation but not for students who do not need it (e.g., Phillips, 1994). A similar concept is echoed in the principle of "differential boost," which holds that an accommodation maintains the validity of the test if it "increases the scores of students with disabilities more than it benefits students without disabilities" (Fuchs & Fuchs, 2001). Accommodations that fulfill this premise are considered to make tests more accurate and fair for those who need the accommodations (Sireci, Scarpati, & Li, 2005) because measurement of the underlying construct does not change because of the accommodation. Similarly, an accommodation that results in improved test scores for all students lacks strong evidence for test validity because the effects are not specific to students with disabilities.

In general, research on the impact of read-aloud accommodations on reading comprehension performance for students with disabilities has yielded positive results, but differences in how read-aloud accommodations are defined have made it difficult to interpret the implications of these study results for practice. A study by Fletcher et al. (2006) tested the effect of a bundled set of accommodations (reading aloud question stems, answer options, and proper nouns and extending the test over 2 days) on a sample of third-grade students in Texas; they found large effects of this accommodation package for students with word-reading disabilities and no effects for students without decoding problems. Fletcher and colleagues hypothesized that the 2-day administration helped because students with reading disabilities experienced fatigue, diminished interest, and burnout during the test because of their difficulties decoding words. Similarly, reading the questions and stems to students reduces demands on word-decoding skills without getting in the way of determining a student's ability to comprehend the passage itself. Reading the proper nouns to the students also reduces their need to rely on decoding skills for these words, potentially decreasing frustration and fatigue, without invalidating the test. Another study (Crawford & Tindal, 2004) examined the impact of a full read-aloud accommodation delivered via video, and it found a differential boost for elementary school students with disabilities on reading comprehension performance versus students without disabilities.

Other recent studies and syntheses lend support to the validity and utility of read-aloud approaches for students with disabilities. For example, Buzick and Stone (2014) and Li (2014) each conducted meta-analyses of studies examining the impact of read-aloud accommodations, and both found that when the outcome was reading comprehension, the accommodations benefited all students but more strongly the students with disabilities. Similarly, Laitusis, Buzick, Stone, Hansen, and Hakkinen (2012) found that among six qualifying studies on elementary school students taking English language arts assessments, all showed a differential boost for students with disabilities who received read-aloud accommodations. The studies that comprised these syntheses varied in their definitions of read-aloud accommodations, with some focusing on full read-aloud (Laitusis, 2010; Meloy, Deville, & Frisbie, 2002), some on read-aloud of proper nouns and/or comprehension stems (Fletcher et al., 2006; Fletcher et al., 2009), and some not specifying (Crawford & Tindal, 2004; Olson & Dirir, 2010). Additionally, some studies examined read-aloud by a human reader (Fletcher et al., 2006; Fletcher et al., 2009; Meloy et al., 2002; Olson & Dirir, 2010), whereas others examined read-aloud via technology (Crawford & Tindal, 2004; Kosciolek & Ysseldyke, 2000; Laitusis, 2010).

Problem and Need

Although recent studies have demonstrated encouraging results for read-aloud accommodations, the topic warrants further investigation. One question that has not been adequately addressed is how students with decoding difficulties respond to the discrete components of testing accommodations that are often bundled. Several possible variations of the read-aloud accommodation have been identified in literature and policy. Some states allow test administration with full read-aloud (encompassing the passages, questions, and answers), whereas others permit question and answer read-aloud only (Kincaid, Thurlow, & Lazarus, 2013). In most cases, these accommodations are also bundled with extended time or pacing assistance. Additionally, while some studies have examined the impact of a full read-aloud accommodation (Laitusis, 2010; Meloy et al, 2002), only one found evidence of differential boost (Laitusis, 2010), and no studies have examined the additive impact of a full read-aloud over other, potentially less resource-intensive accommodations, such as reading aloud questions stems and answer options and/or reading aloud individual words.

In the current study, our aim was to build on the literature that examines read-aloud support by evaluating each aspect of such support. Our study unbundled the different types of accommodations examined by Fletcher et al. (2006) and included a full read-aloud component, to determine whether and in what way adding successively more support leads to significant benefit while maintaining the validity of the assessment.

Research Questions

The current study examined four research questions:

Research Question 1: What is the benefit of providing a read-aloud of question stems and answer options on a reading comprehension test for students with word-reading difficulties versus students with no word-reading difficulties?

Research Question 2: What is the benefit of providing a read-aloud of question stems, answer options, and proper nouns on a reading comprehension test for students with word-reading difficulties versus students with no word-reading difficulties?

Research Question 3: What is the benefit of providing a full read-aloud on a reading comprehension test for students with word-reading difficulties versus students with no word-reading difficulties?

Research Question 4: For students with word-reading difficulties, what is the added benefit of each level of increased read-aloud support?

Research Questions 1–3 were designed to test the aforementioned concepts of “interaction hypothesis” and “differential boost,” which posit that test accommodations should benefit students with disabilities but not students without disabilities (interaction hypothesis; Phillips, 1994) or that they should benefit students with disabilities significantly more than they benefit students without disabilities (differential boost; Fuchs & Fuchs, 2001). Therefore, our study evaluated the effect of various accommodations on a sample of students with and without word-reading difficulties to detect differences in the effects of the accommodations between these two groups. This design allowed us to be more confident in the validity of the test in cases where the accommodation resulted in improved test scores for the sample of students with word-reading difficulties but not for those without word-reading difficulties.

Method

Recruitment Process

We worked directly with staff at the Ohio Department of Education to identify schools interested in participating in the study.[1] We initially targeted large districts and districts that had expressed to the state education agency some interest in reading disabilities and accommodations. Twelve districts agreed to support the study by communicating relevant information and soliciting interest from their schools. We also conducted outreach through a variety of state networks to recruit additional schools outside of these districts. The schools that composed our final sample were generally clustered around three large cities in the state. Each participating school was provided with a packet of information about the purpose of the study, procedures for identifying students, and the study instrument, with consent forms for all participants. Schools also received a monetary incentive for participation in the study.

School principals, or another appropriate point of contact, were asked to identify current fourth graders who were potentially eligible for the study, including students who had or might have a reading disability (e.g., had an IEP with a reading-related disability or goal, or had been referred for special education due to reading difficulty), students who scored at a “basic” level or below on the third-grade reading assessment, and students who did not have any identified disability other than a learning disability (e.g., a student who had been identified with emotional disturbance, autism, hearing impairment, or vision impairment). In addition, principals were asked to identify an equal number of students (as meeting the preceding criteria) considered to be “average readers” (i.e., students with no disabilities and scoring at approximately the 50th percentile in reading). All recruitment, screening, and data collection procedures were approved by the appropriate Institutional Review Boards.

Student Screening Process

The research team hired and trained testing personnel to administer the screening tools to all identified students whose parents provided consent.[2] We used the following two instruments to determine which students qualified for inclusion in the study sample:

- *Woodcock-Johnson III Diagnostic Reading Battery–Basic Reading Cluster* (Letter-Word Identification and Word Attack; Woodcock, Mather, & Schrank, 2004)
- *Test of Word Reading Efficiency 2* (Sight Word Efficiency and Phonemic Decoding Efficiency; Torgesen, Wagner, & Rashotte, 2012)

These measures were administered individually, and sessions were audio recorded. For the *Woodcock-Johnson III Diagnostic Reading Battery* Letter-Word Identification subtest, students read aloud from lists of increasingly difficult real words. For the Word Attack subtest, students read aloud from lists of increasingly difficult nonwords. For the *Test of Word Reading Efficiency 2* Sight Word Efficiency subtest, students had 45 s to read aloud from a list as many increasingly difficult words as they could. For the Phonemic Decoding Efficiency subtest, students had 45 s to read aloud from a list as many increasingly difficult nonwords as they could.

Students scoring <26th percentile on three of the four subtests qualified for inclusion in the study sample as students with word-reading difficulties. Students scoring between the 34th and 76th percentiles on three of four subtests qualified for inclusion in the study sample as average readers. All other students were excluded from participation in the study. Of 701 students screened, 346 (49%) qualified for participation in the study. The screening criteria that we used and the percentage of students who were eligible based on these criteria were consistent with the procedures used in Fletcher et al. (2006).

Sample

Our original sample included 224 fourth-grade students in 46 schools across 26 districts in the state of Ohio, all of whom met the screening criteria described in the previous section (136 met the criteria for decoding difficulty and 88 met the criteria for average reader). These 46 schools were randomly assigned to four conditions, which varied with respect to the order in which students in that school

Table 1. Sample Sizes.

Group	Sample, *n*		
	Data collection	Initial analytic	Final analytic
Average readers	88	82	49
Reading difficulty	136	125	96
Total	224	207	145

experienced the different types of accommodation (for details on this randomization process, see Study Design and Data Collection Procedures). The assignment procedure was done in such a way to ensure equal numbers of students in each condition. After data collection but prior to analysis, we removed cases where students were absent for all of the testing session, which reduced the sample size to 207 (125 with reading difficulty and 82 average readers). Next, we conducted sensitivity analyses (described in the Analyses section) that removed students who had incomplete data or who indicated familiarity with the testing passages; thus, the sample size was further reduced to 145 (96 with reading difficulty and 49 average readers). Table 1 shows the number of students with and without word-reading difficulties for the data collection sample, initial analytic sample, and final analytic sample. Despite these smaller sample sizes, we detected meaningful statistically significant effects.

Measures

Our primary outcome measure was reading comprehension as assessed on the state reading measure. Passages from the reading portion of the fourth-grade state reading measure were initially drawn from a set of 18 released passages and associated items used in state assessments from 2005 to 2011. The copyright permissions expired for 5 passages, and 2 passages were poems; therefore, 11 passages were available to be used to form tests. These passages had between five and seven multiple-choice questions, 400 to 700 words, and a Flesch reading ease score of 70 to 98.

Our goal was to create four test booklets consisting of two passages each. We sought to create booklets that were similar to each other in difficulty, number of multiple-choice items, and length. For all 55 possible combinations of test passages into booklets, booklet difficulties were computed through item response theory test information function, total number of items, as well as measures related to passage length and readability, such as number of words and Flesch reading ease score. From the resulting test configurations, we chose four booklets that had the same test difficulty with a similar number of items (between 11 and 14), as well as equivalence in passage length and readability.

We used the item response theory scaling from the existing state assessment item parameters to estimate student scores on the item blocks via Rasch scoring. To compute the scale scores, the item difficulties were used for the released items and transformation coefficients for converting theta or *z* scores into operational Grade 4 state assessment reading scale scores. We assessed the extent to which items no longer fit the Rasch model, using the Rasch INFIT and OUTFIT criteria. These final scale scores were used as outcomes in our analyses.[3]

Study Design and Data Collection Procedures

Thirteen trained test administrators, most of whom were school psychology interns, screened and tested students in the study sample. The test administrator administered the study instrument in a group setting, with the number of students per school ranging from 1 to 10. The 2-hr testing sessions[4] were conducted over 2 consecutive days. The same eight passages were used with all students, and the order of passages and accommodation conditions was randomized.

We used a Latin square design for the study so that the order in which we presented support levels would counterbalance the potential order effects that we might see related to providing different levels of support. A Latin square design assigns each treatment condition to each row and column of a J × J matrix corresponding to treatment trials and participant blocks. Thus, each treatment is administered to each participant, and the order is both randomly assigned to participants and balanced across participants (Box, Hunter, & Hunter, 1978). Students completed four test sessions over 2 consecutive days (two sessions on each day), operationalized as follows:

No accommodations: Students read the passage, questions, and answer choices silently to themselves, and each student had up to 1 hr to complete two blocks.

Questions and answer choices read aloud: Students read the passage silently; then, the test administrator read aloud the first question and answer choices. After all students had answered the question, the administrator read aloud the next question and answer choices and so on for all questions.

Questions, answer choices, and proper nouns read aloud: Before students read the passage silently to themselves, the test administrator read aloud a list of proper nouns that would appear in the passage, which was printed separately for the students so that they could follow along. Students then read silently to themselves. When all students had finished reading, the same procedures for reading questions and answers were followed as indicated.

Full read-aloud: The test administrator read aloud the passage with all questions and answer choices. Similar to the previous accommodations, for each question, the administrator waited until all students answered before reading aloud the next question.

Table 2. Illustration of Latin Square Design, With Sample Sizes by Counterbalancing Condition.

Test session	Condition 1[a]	Condition 2[b]	Condition 3[c]	Condition 4[d]
Day 1				
Session 1 (S1)	No accommodations	Read-aloud of question stems and answer options	Read-aloud of question stems, answer options, and proper nouns	Full read-aloud
Session 2 (S2)	Read-aloud of question stems and answer options	Read-aloud of question stems, answer options, and proper nouns	Full read-aloud	No accommodations
Day 2				
Session 1 (S3)	Read-aloud of question stems, answer options, and proper nouns	Full read-aloud	No accommodations	Read-aloud of question stems and answer options
Session 2 (S4)	Full read-aloud	No accommodations	Read-aloud of question stems and answer options	Read-aloud of question stems, answer options, and proper nouns

Note. DC = data collection sample; IA = initial analytic sample; FA = final analytic sample.
[a]DC, n = 51; IA, n = 48; FA, n = 34. [b]DC, n = 55; IA, n = 52; FA, n = 40. [c]DC, n = 53; IA, n = 44; FA, n = 25. [d]DC, n = 65; IA, n = 63; FA, n = 46.

For each session, students received a test booklet comprised of two "blocks," and they completed two sessions each on two consecutive days, for a total of four sessions (or eight blocks). Based on the Latin square design (see Table 2), the order of administration for the accommodations was spiraled across participating schools to account for any order effects. In addition to controlling for the order effects, booklet effects were controlled by combining each test booklet with every set of reading accommodations and randomly presenting them under the Latin square design in a way that resulted in each possible combination being administered a similar number of times to a similar number of students.

Analyses

Impact analyses. The study used a three-level hierarchical linear model, in which observations from the accommodation conditions were nested within students and students were nested within schools, to assess whether there were significant differences between students with word-reading difficulties and average readers in the effect of each accommodation relative to no accommodation. In addition to the main effects of the accommodation conditions and reading difficulty indicator, the model testing Research Questions 1 to 3 included interaction terms between the reading difficulty indicator and accommodation conditions to test the interaction hypothesis. The analysis model was of the following form:

$$\begin{aligned} Y_{ijk} = {} & \beta_{0j} + \beta_1 RD_{ij} + \beta_2 Questions + \\ & \beta_3 Proper\_Nouns + \beta_4 Full + \\ & \beta_5 RD * Questions + \beta_6 RD * Proper\_Nouns + \\ & \beta_7 RD * Full + \delta i + \vartheta j + \varepsilon_{ijk}, \end{aligned} \tag{1}$$

where Y_{ijk} represents the reading scale score for student i in school j under accommodation condition k, RD refers to whether the student has reading difficulty, *Questions* refers to the read-aloud of question stems and answers, *Proper_Nouns* is for the read-aloud of proper nouns accommodation, and *Full* refers to the full read-aloud accommodation. The analysis also included student random effects, δi, and school random effects, ϑj, to account for the data structure where observations were clustered within students and within schools.

To answer Research Question 4, we estimated models using only the data from students with word-reading difficulties. This model examined the additional effect of each read-aloud accommodation and is based on the following form:

$$\begin{aligned} Y_{ijk} = {} & \beta_{0j} + \beta_2 Questions + \beta_3 Proper\_Nouns + \\ & \beta_4 Full + \delta i + \vartheta j + \varepsilon_{ijk}, \end{aligned} \tag{2}$$

where *Questions, Proper_Nouns*, and *Full*, like before, differentiated performance with respect to no accommodation.

Sensitivity analyses. We also conducted sensitivity analyses to examine whether the results were sensitive to the characteristics of the analysis sample. First, we replicated all the analyses after removing the cases that had any incomplete outcome data across accommodations to test whether the results obtained from the main analyses were sensitive to exclusion of incomplete data.[5]

Second, we replicated the analyses after removing cases from eight schools in which test administrators reported to us that the students indicated familiarity with the reading passages. We hypothesized that such familiarity would

Table 3. Average Screening Test Results for Students With Reading Difficulty With and Without IEPs.

Measure	*M* (*SD*)		Different
	With IEP[a]	Without IEP[b]	
Woodcock-Johnson			
Letter-Word Identification	77.94 (8.68)	84.75 (3.78)	Yes
Word Attack percentile	85.31 (8.26)	87.82 (4.24)	No
Test of Word Reading Efficiency 2			
Sight Word Efficiency	70.47 (10.49)	80.93 (5.47)	Yes
Phonemic Decoding Efficiency	67.54 (8.08)	71.50 (7.48)	Yes

Note. "Yes" indicates that the two groups are significantly different from each other at $p < .05$. IEP = individualized education program.
[a]$n = 68$. [b]$n = 28$.

increase the possibility of students answering the questions correctly regardless of the read-aloud accommodations, therefore biasing the results. Note that the team did not systematically explore the degree to which students were exposed to the passages in all schools in the study. Therefore, it is possible that prior exposure to the passages was an issue in other schools as well and that the students simply did not report it. Given this possibility, results should be interpreted with caution.

Supplemental exploratory analyses. As a final exploratory analysis, we examined whether having an IEP was associated with benefit of the accommodations. Because we defined our sample of students with reading difficulties as those who fell below a predetermined cutoff on a set of screening instruments, our sample did not necessarily include only students on IEPs. Sixty-six percent of our final sample of students with word-reading difficulties had IEPs. Descriptive analyses of the data showed that among our sample of students with word-reading difficulties, those with IEPs had systematically lower screening scores than did those without IEPs (see Table 3).

Results

In this study, we examined the differential benefits of providing three read-aloud accommodations: read-aloud of question stems and answer options; read-aloud of proper nouns, question stems, and answer options; and full read-aloud (Research Questions 1–3). We also investigated the added benefit of each level of increased read-aloud support for students with word-reading difficulties (Research Question 4). We ran the analyses after standardizing the outcomes variables (i.e., scale scores) so that the results could be interpreted as the effect sizes. In this section, we report the results for the parameters of interest: either the parameter for the interaction term or the parameter that shows the difference between a reading accommodation and no accommodation at all.

Based on our initial analytic sample (n = 207), results showed that even though the differential effect sizes for the read-aloud accommodations (i.e., question stems and answer options; effect size = 0.28, p = .144), the proper nouns accommodation (effect size = 0.35, p = .067), and full read-aloud accommodation (effect size = 0.25, p = .181) were of considerable magnitude, they were not statistically significant. However, according to our final analytic sample (n = 145), results differed. For all three read-aloud accommodations, there was a statistically significant differential effect favoring the group of students with word-reading difficulties. The effect sizes were 0.49 (p = .035), 0.55 (p = .017), and 0.50 (p = .031), respectively, for the following accommodations: question stems and answer options read aloud; question stems, answer options and proper nouns read aloud; and full read-aloud (see Table 4). We compared the initial analytical sample to the final analytical sample in terms of screening measures. Table 5 shows that the differences between the two groups on these measures were ≤2 points and not statistically significant, suggesting that excluding these cases does not significantly alter the composition of the sample. Descriptive analyses showed that for both groups of students, the proper nouns condition was the least beneficial. In fact, the average readers in our sample performed worse with the proper nouns accommodation than with no accommodation (see Figure 1). This accounted for the relatively large differential effect (effect size = 0.55, p = .017) that was found for the proper noun accommodation.

For the final research question, we focused on only students with word-reading difficulties. Using the initial analytic sample, we found that the accommodation of question stems and answer options (effect size = 0.26, p = .008) and the full read-aloud accommodation (effect size = 0.37, p < .001) had statistically significant effects on reading performance when compared with no accommodation. With our final analytic sample, we found similar results (see Table 6). There was a statistically significant effect for the accommodation of question stems and answer options (effect size = 0.24, p = .028) and the full read-aloud accommodation (effect size = 0.30, p = .006) when compared with no

Table 4. Effect of Reading Accommodations on Students' Reading Scores in Effect Size Units (Full Hierarchical Linear Model Results).

	Effect size (*SE*)		
Accommodation	Initial analytical sample[a]	Final analytical sample[b]	After removing non-IEP students[c]
Disability	–0.885***	–1.173***	–1.213***
	(0.134)	(0.166)	(0.177)
Questions	–0.0120	–0.248	–0.381*
	(0.145)	(0.188)	(0.191)
Proper nouns	–0.222	–0.397*	–0.238
	(0.148)	(0.188)	(0.191)
Full read-aloud	0.118	–0.199	–0.191
	(0.147)	(0.188)	(0.191)
Disability × Questions	0.275	0.490*	0.748**
	(0.188)	(0.232)	(0.250)
Disability × Proper Nouns	0.347	0.552*	0.473
	(0.190)	(0.232)	(0.250)
Disability × Full Read-Aloud	0.253	0.500*	0.593*
	(0.189)	(0.232)	(0.250)
Constant	0.435***	0.710***	0.617***
	(0.106)	(0.136)	(0.135)

Note. IEP = individualized education program.
[a]n = 207. [b]n = 145. [c]n = 117.
*$p < .05$. **$p < .01$. ***$p < .001$.

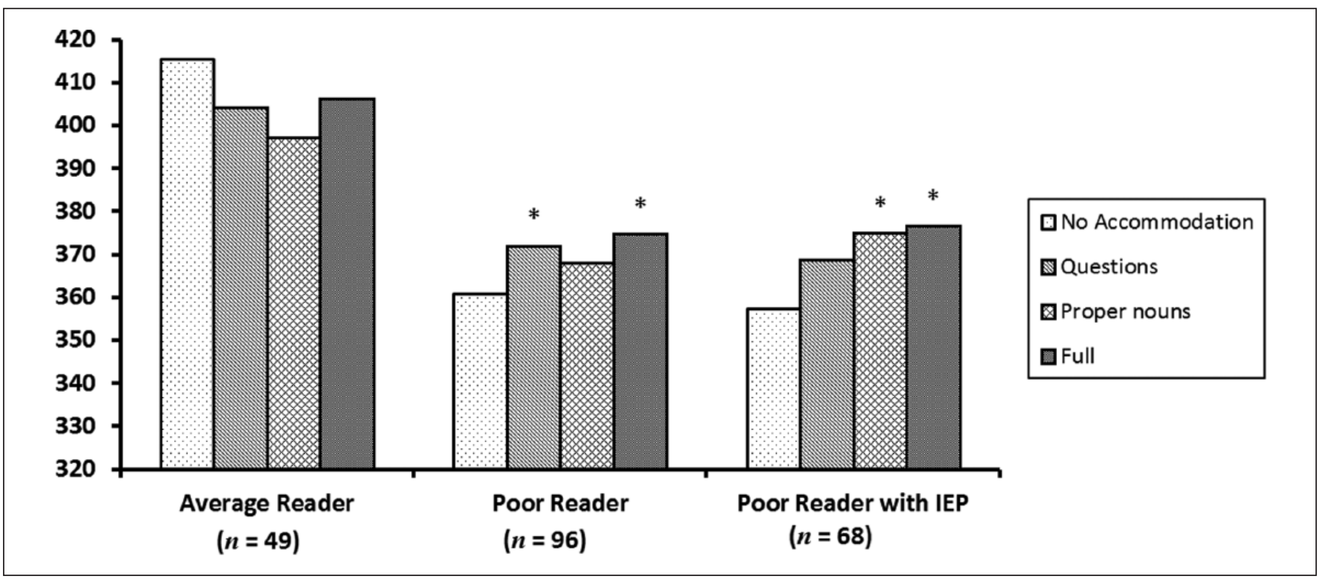

Figure 1. Average scale scores by reading accommodation notes. *Note*. These statistics are based on data after removal of students with any missing data and students in schools reporting familiarity with the reading passages. *Different from "No Accommodation" at $p < .05$.

accommodation. The effect for the proper nouns accommodation was not statistically significant. Furthermore, the effects for the accommodation of question stems and answer options and the full read-aloud accommodation were not statistically significantly different from each other, suggesting that for this sample, both accommodations had similar benefits in terms of impact on performance.

Results for Students With IEPs

One of the issues that we examined through supplemental exploratory analysis was the degree to which the results might vary if the group of students with word-reading difficulties included only students who had IEPs; this group (n = 68) accounted for 66% of the students with word-reading

Table 5. Comparison of the Initial Analytical Sample With Final Analytical Sample.

	Initial analytical sample[a]		Final analytical sample[b]		
Measure	*M*	*SE*	*M*	*SE*	Different
Woodcock-Johnson					
Letter-Word Identification	89.13	0.88	87.30	1.05	No
Word Attack percentile	93.94	0.77	92.16	0.92	No
Test of Word Reading Efficiency 2					
Sight Word Efficiency	84.45	1.10	83.03	1.34	No
Phonemic Decoding Efficiency	80.86	1.12	78.39	1.32	No

Note. "Yes" indicates that the final analytical sample is different from the initial analytical sample at $p < .05$.
[a]$n = 207$. [b]$n = 145$.

Table 6. Effect of Reading Accommodations on Reading Scores of Students With Reading Difficulties in Effect Size Units.

	Effect size (*SE*)		
Accommodation	Initial analytical sample[a]	Final analytical sample[b]	After removing non-IEP students[c]
Questions	0.263**	0.242*	0.235
	(0.0987)	(0.110)	(0.130)
Proper nouns	0.124	0.155	0.366**
	(0.098)	(0.110)	(0.130)
Full read aloud	0.369***	0.300**	0.401**
	(0.098)	(0.110)	(0.130)
Constant	–0.437***	–0.449***	–0.588***
	(0.0772)	(0.0854)	(0.0989)

Note. The results represent the size (and standard errors) of the effect of accommodations relative to no accommodations for students with reading difficulties from the regression models described earlier. IEP = individualized education program.
[a]$n = 125$. [b]$n = 96$. [c]$n = 68$.
*$p < .05$. **$p < .01$. ***$p < .001$.

difficulties in our sample and had systematically lower screening scores than students who did not have IEPs (see Table 3). By analyzing the data with a subsample of students with word-reading difficulties who had IEPs, we tested whether the results would be different for the students at the lowest end of the range for screening scores.

Our results were indeed sensitive to sample restriction.[6] While the pattern of differential effects based on the restricted sample was similar to the pattern based on the full sample, the differential effect based on the restricted sample for the condition where question stems and answer options were read aloud was not statistically significant,[7] and the differential effect size for the proper nouns condition was notably larger in magnitude than that based on the full sample (0.75, $p = .003$, vs. 0.55, $p = .017$; see Table 4 and Figure 1). Also, the effect size for the proper nouns condition was larger than that for the full sample, and it was statistically significant (when it had not been before). Additionally, the effect size for the full read-aloud was larger than it had been for the full sample (Table 6). These findings indicate that the proper nouns and full read-aloud conditions may hold greater benefits for students with identified disabilities.

Discussion

This study investigated four research questions about the impact of read-aloud accommodations. The first three, respectively, examined the differential benefit of providing (a) a read-aloud of question stems and answer options; (b) a read-aloud of question stems, answer options, and proper nouns; and (c) a full read-aloud on a reading comprehension test for students with word-reading difficulties versus students with no word-reading difficulties. The fourth research question examined the added benefit of each level of increased read-aloud support for students with word reading difficulties.

Our results showed that all three read-aloud accommodations benefited students with reading difficulties more than they did students without. For students with word-reading difficulties only, the results showed statistically significant effects for two of the accommodations—read-aloud of question stems and answer options and full read-aloud—as compared with no accommodation at all. The proper nouns accommodation did not show a statistically significant impact on performance for students with word-reading difficulties when compared with no accommodation at all. Furthermore, there was no evidence of an additive benefit for each increased level of support. In other words, the results showed that the full read-aloud was equally beneficial to reading aloud question stems and answer options.

Additional analyses on a restricted sample of students who had IEPs (and who had the lowest word-reading screening scores) showed some differences from the main sample results. Specifically, there was a stronger impact of the proper nouns and full read-aloud accommodations as compared with no accommodation at all than there had been for the full sample. The differential impact of the proper nouns accommodation was also much larger than it was for the full sample.

These findings extend the previous literature on read-aloud accommodations by examining the impact of discrete components of typical bundled read-aloud packages and their impact for students with different degrees of decoding difficulty. Two points are important to note. First, our study demonstrated the full read-aloud accommodation to be beneficial for students with word-reading difficulties, building on previous findings by Laitusis (2010). However, the effect for the full read-aloud versus no accommodation at all was not statistically significantly different from the effect of reading aloud just the question stems and answer options as compared with no accommodation at all. For our restricted sample of students with IEPs (who had the lowest word-reading skills), the effect for the full read-aloud was greater than it was for the full sample. This is important, as providing a full read-aloud option can be controversial; some researchers have argued that when a full read-aloud is implemented, the construct under evaluation shifts from reading comprehension to listening comprehension (Crawford & Tindal, 2004). Furthermore, implementing a full read-aloud can be resource intensive and thus perhaps not realistic to implement in practice. However, as our study has shown, it has benefits for students with very low decoding skills. For districts that are weighing whether to devote resources to implementing full read-aloud accommodations, it is helpful to know the students for which it is likely to yield the greatest benefit. In other words, for students whose decoding skills are on the very low end of the spectrum, implementing a full read-aloud accommodation may be the most beneficial, whereas for students with less intensive needs, read-aloud of question stems and answer options (which is less costly) may be sufficient.

Second, our study unbundled the impact of a read-aloud of proper nouns from a read-aloud of question stems and answer options, extending findings from Fletcher et al. (2006). We found that the proper noun accommodation had varying effects on the performance of different groups of students. For average readers, it seemed to have a negative effect. For the overall sample of students with word-reading difficulties, the proper nouns accommodation had a positive but not statistically significant effect; for the subset of students with IEPs (who had the lowest word-reading skills), this accommodation had a stronger and statistically significant effect. These findings suggest that districts may want to offer support to target specific words in the text for only students falling below a certain threshold because it appears that reading aloud of proper nouns is helpful especially to students with very poor word-reading skills. Although it was beyond the scope of this study, further study may be warranted to investigate whether it could be helpful to offer support for reading additional individual words in the text for this set of students. Given the increased use of technology to deliver assessments, this type of individualized accommodation may be especially feasible and desirable.

Limitations and Future Research

Although this study extended previous work conducted by Fletcher and colleagues, it was not a direct replication of the 2006 study. The sample of students in Fletcher et al. (2006) had overall lower reading ability, and all had IEPs. Despite these differences, it is notable that the findings are generally aligned with Fletcher's study and with other recent studies on read-aloud accommodations (e.g., Buzick & Stone 2014; Crawford & Tindal, 2004; Lai & Berkeley, 2012; Laitusis, 2010; Laitusis et al., 2012; Li, 2014). The current study extends our understanding of testing accommodations by isolating their independent effects and adding an examination of a full read-aloud accommodation. One potential limitation, however, is that we do not know whether students in our sample received similar accommodations during their instructional time. Exposure to similar accommodations in the classroom (or lack thereof) could have an effect on the degree to which the accommodations benefited them in the testing environment. Another limitation relates to the fact that the assessment was group administered, raising the possibility that students' answers could have been affected by the behavior of their peers in their groups. Additionally, although we excluded students from the final sample who had divulged prior exposure to the test passage, it is possible that students in other schools had experienced the same but did not report this to the test administrators.

From these findings and limitations, we recommend research that examines the relationship between accommodations offered in the classroom and those offered on tests; specifically, more information is needed on the degree to which alignment in accommodation practices supports better outcomes. This is especially important given provisions in the 2004 amendments to the Individuals with Disabilities Education Act that require states and districts to ensure that students with print disabilities are provided with accessible instructional materials. Many publishers now offer digital materials to support students who have difficulty accessing print through reading, and as a result, schools have many more options for providing read-aloud accommodations during instruction.

Additionally, we recommend that research take into account the growing use of computer-based platforms to deliver large-scale assessments and offer access to print during instruction. For example, computer-based assessments may be better able to individualize accommodations so that students with word-reading disabilities can self-select the words for which they need decoding support, as opposed to receiving a full read-aloud or partial support for only a select group of words, such as proper nouns. Computer-based assessments also remove the need for a human reader to read aloud, which could increase or decrease their validity. Future research might also assess the impact of using a human reader versus an electronic reader.

However, regardless of the platform, this study reinforces the existing literature that shows the need for read-aloud accommodation support during large-scale assessments so that comprehension skills of students with reading disabilities can be assessed fairly. States and the large-scale assessment consortia should continue to permit read-aloud accommodations, including full read-aloud, where financially feasible. Finally, IEP teams should continue to make accommodation decisions on a case-by-case basis, taking into consideration the unique needs of each student with a disability.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by Grant R324A120110 from the U.S. Department of Education, Institute of Education Sciences, to American Institutes for Research.

Notes

1. We conducted this study in collaboration with the state of the Ohio because of its recent adoption of a "Third Grade Reading Guarantee" policy and its interest in learning about strategies to support improved reading outcomes for students with disabilities.
2. In some cases, a school staff member conducted the screening. This staff member was experienced in administering the instruments and was provided compensation for taking on this responsibility.
3. However, we replicated our analyses using the percentage correct and scale score that uses all the items to ensure that the results are not sensitive to the computation of scale scores.
4. The study was not designed to be timed at all, and the 2 hr allotted proved to be more than enough time for every student to complete the test.
5. The students with incomplete data were removed from the data to guard against any confounding factor that might have prevented these students from completing the assessment under all four conditions. However, only five students had at least one incomplete outcome data point. The results with and without these five students were very similar, and the inferences made were identical.
6. Even though removing non-IEP students greatly reduced the sample size and statistical power, the impact of some accommodations became large enough to be detected as significant.
7. The insignificant result is most likely due to the reduced sample, because the effect size for this accommodation did not change much when the non-IEP students were removed from the analysis.

References

Box, G. E. P., Hunter, J. S., & Hunter, W. G. (1978). *Statistics for experimenters: An introduction to design, data analysis, and model building*. New York, NY: Wiley.

Buzick, H., & Stone, E. (2014). A meta-analysis of research on the read-aloud accommodation. *Educational Measurement: Issues and Practice*, *33*(3), 17–30.

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

Fletcher, J. M., Francis, D. J., Boudousquie, A., Copeland, K., Young, V., Kalinowski, S., & Vaughn, S. (2006). Effects of accommodations on high-stakes testing for students with reading disabilities. *Exceptional Children*, *72*(2), 136–152.

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

Fuchs, L. S., & Fuchs, D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research & Practice*, *16*(3), 174–181.

Fuchs, L., Fuchs, D., & Capizzi, A. (2005). Identifying appropriate test accommodations for students with learning disabilities. *Focus on Exceptional Children*, *37*, 1–8.

Hodgson, J. R., Lazarus, S. S., & Thurlow, M. L. (2010). *Characteristics of states' alternate assessments based on modified academic achievement standards in 2009–2010* (Synthesis Report No. 80). Minneapolis, MN: National Center on Educational Outcomes.

International Dyslexia Association. (2008). *Dyslexia basics*. Retrieved from www.interdys.org/ewebeditpro5/upload/Basics_Fact_Sheet_5-08-08.pdf

Jamgochian, E. M., & Ketterlin-Geller, L. R. (2015). The 2% transition: Supporting access to state assessments for students with disabilities. *Teaching Exceptional Children*, *48*(1), 28–35.

Kincaid, A., Thurlow, M., & Lazarus, S. (2013). *2012 state policies for accommodations used to deliver assessments orally*. Minneapolis, MN: National Center on Educational Outcomes.

Kosciolek, S., & Ysseldyke, J. E. (2000). *Effects of a reading accommodation on the validity of a reading test* (Technical Report No. 28). Minneapolis, MN: National Center on

Educational Outcomes. Retrieved from https://nceo.info/Resources/publications/OnlinePubs/Technical28.htm

Lai, S. A., & Berkeley, S. (2012). High-stakes test accommodations research and practice. *Learning Disability Quarterly*, *35*(3), 158–169.

Laitusis, C. C. (2008). State reading assessments and the inclusion of students with dyslexia. *Perspectives on Language and Literacy*, *34*(4), 31–33.

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

Laitusis, C. C., Buzick, H., Stone, E., Hansen, E., & Hakkinen, M. (2012). *Literature review of testing accommodations and accessibility tools for students with disabilities*. Princeton, NJ: Educational Testing Service.

Lazarus, S. S., Cormier, D. C., & Thurlow, M. L. (2011). States' accommodations policies and development of alternate assessments based on modified achievement standards: A discriminant analysis. *Remedial and Special Education*, *32*(4), 301–308.

Lazarus, S. S., Thurlow, M. L., Eisenbraun, K. D., Lail, K. E., Matchett, D. L., & Quenemoen, M. (2006). *State accommodations policies: Implications for the assessment of reading*. Minneapolis, MN: Partnership for Accessible Reading Assessment.

Li, H. (2014). The effects of read-aloud accommodations for students with and without disabilities: A meta-analysis. *Educational Measurement: Issues and Practice*, *33*(3), 3–16.

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.

Olson, J. F., & Dirir, M. D. (2010). *Technical report for studies of the validity of test results for test accommodations*. Washington, DC: Council of Chief State School Officers.

Partnership for Assessment of Readiness for College and Careers. (2015). *PARCC accessibility features and accommodations manual 2015–2016*. Washington, DC: PARCC Assessment Consortia.

Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education*, *7*(2), 93–120.

Shaywitz, S. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level*. New York, NY: Knopf.

Sireci, S. G. (2004). *Validity issues in accommodating NAEP reading tests* (Center for Educational Assessment Research Report No. 515). Amherst, MA: University of Massachusetts Amherst.

Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, *75*(4), 457–490.

Smarter Balanced Assessment Consortium. (2015). *Usability, accessibility, and accommodations guidelines*. Retrieved from www.smarterbalanced.org/wp-content/uploads/2015/09/Usability-Accessibility-Accomodations-Guidelines.pdf

Thurlow, M. L., Moen, R. E., Liu, K. K., Scullin, S., Hausmann, K. E., & Shyyan, V. (2009). *Disabilities and reading: Understanding the effects of disabilities and their relationship to reading instruction and assessment*. Minneapolis, MN: Partnership for Accessible Reading Assessment.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency–Second Education (TOWRE-2)*. Austin, TX: Pro-Ed.

U.S. Department of Education. (2013, August 23). *Department of Education proposes to eliminate "2 percent rule" in assessing students with disabilities* [Press release]. Retrieved from www.ed.gov/news/press-releases/department-education-proposes-eliminate-2-percent-rule-assessing-students-disabilities

Woodcock, R. W., Mather, N., & Schrank, F. A. (2004). *Woodcock-Johnson III Diagnostic Reading Battery*. Rolling Meadows, IL: Riverside.