

Table of Contents

Introduction.....	1
The Nature of Reading.....	1
Reading and Disabilities	2
Issues in Measurement	3
Accommodations in Reading Assessment	4
Partial Auditory Accommodation.....	6
Methods	8
Participants	8
Measures	8
Materials	9
Procedures.....	9
Results.....	10
Effects of the Reading Pen.....	10
CBM-R and Maze.....	11
Student Evaluations of the Reading Pen	12
Discussion	14
Conclusion	15
Limitations and Future Research.....	16
References	17
Appendix A. The Reading Pen Instruction Packet.....	23
Appendix B. Practice Paper.....	25
Appendix C.	27

Introduction

Students attending public schools across the U.S. participate in annual state assessments designed to provide a picture of what students know and are able to do. All students participate in these assessments, including students with disabilities. Yet prior to the past decade, the rates of participation of students with disabilities in large scale assessments were as low as 10% (Shriner, Spande, & Thurlow, 1994). Participation in state assessments now is required by federal law, and the number of students with disabilities who participate in annual testing for accountability has increased dramatically (Thurlow, Quenemoen, Altman, & Cuthbert, 2008), to the point that nearly all states report at least 95% of students with disabilities participating in state assessments (Thurlow, Moen, & Wiley, 2005). With this increased participation, there is increased interest in ensuring that students are measured accurately, and that knowledge and skills that students do have are not obscured by the nature of the assessment itself.

Despite strong recommendations that state assessments should be universally designed (Thompson, Johnstone, & Thurlow, 2002; Thompson, Thurlow, & Malouf, 2004) and accessible to the widest range of students (IDEA, 2004; NCLB, 2001), researchers and advocates continue to question whether these goals have been met (Thurlow, in press; Thurlow, Laitusis, et al., 2009). It has been suggested that characteristics associated with some disabilities may block a student's access to the content of the test, especially in the case of reading assessments (Thurlow, Moen, Liu, et al., 2009), with the result being that students are not able to show their knowledge and skills simply because the assessment itself has created a barrier to doing so.

Concern about whether assessments are providing the best indication of students' knowledge and skills has arisen, in part, because researchers have found that the state assessment performance of students with disabilities, on average, consistently is lower than that of students without disabilities (Abedi, Leon, & Mirocha, 2003; Darling-Hammond, 2003; Thurlow, Bremer, & Albus, 2008). The extent to which this lower average performance is due to lower levels of knowledge and skills rather than barriers in the assessment itself is not yet known, although research is exploring this issue (Dolan, Hall, Banerjee, Chun, & Strangman, 2005; Johnstone, 2003; Moen, et al., 2009).

The Nature of Reading

The ability to read can be understood as the use of several skills that function together to support a reader's ability to acquire and make use of information represented symbolically. The National Reading Panel (2000) identified five subskills that are necessary for proficient reading: phonemic awareness, phonics, fluency, vocabulary, and comprehension. These subskills can be separated into two main groups: word level skills and broader text level skills. Word level skills include phonemic awareness and phonics, and broader text level skills include fluency, vocabulary, and comprehension.

Stahl and Hiebert (2005) described reading fluency as a product of accuracy, rate, and

prosody. Accuracy in reading is a function of a person's skill in decoding unfamiliar words, and automatically recognizing familiar words. Word recognition also supports the rate at which reading can occur. Prosody allows for reading that reflects conventions of the spoken language such as the pause between two sentences representing a separation between two ideas. For beginning readers, much instructional emphasis is placed on word level skills or decoding. As students grow and the complexity of tasks increases, the importance of decoding is surpassed by automatic word recognition and correct prosody, which together create fluency.

Aside from the foundational skills involved in reading fluency, reading comprehension also relies on linguistic and cognitive skills. Strong readers are expected to be able to recall, integrate, interpret, and evaluate various qualities of information read from text. They have vocabularies large enough to both recognize specific words and understand word meanings in context. They have a foundation of fluency that supports comprehension. They are able to sustain a pace in reading that allows information to be added to a developing idea before prior information is forgotten. These skills all operate at the same time to produce comprehension (National Assessment Governing Board, 2008).

Because of the multifaceted nature of the process of reading comprehension, there are a number of points at which problems can occur. For instance, some students might have adequate skill in word-level decoding but perform poorly on a reading test due to deficits in prosody. Other students could have adequate prosody, but might perform poorly on a reading test due to deficits in decoding or word recognition. Some students may do poorly on reading tests due to diminished motivation or self efficacy. Recognizing the diversity of students and the multifaceted nature of reading suggests that it is a challenge to accurately measure students' reading comprehension.

Reading and Disabilities

Many students in U.S. schools receive special education services for disabilities that affect reading. The extent to which we understand how specific disabilities affect performance is influenced by how well we understand the disability. Some disabilities are directly related to the constructs and skills that are being measured by an assessment. This is often the case for tests of reading, where students with learning disabilities in reading have specific disability-related limitations in what the test is intending to measure.

Educators and researchers have noted the heterogeneity in characteristics of children with reading disabilities (Fuchs & Fuchs, 1999; Morris et al., 1998). This heterogeneity contributes to the difficulty in determining the extent to which disabilities that affect reading are related to the construct being measured in an assessment. Although the methods used to identify students with reading disabilities are somewhat standardized, student instructional needs vary, as do ways in which students might best access reading tests. Whether determining proper methods of instruction, intervention, or accommodation, difficulties arise in the identification of individual student needs.

The ways in which reading problems are defined and identified often depend on the way comprehension is assessed and properties of the instruments used for assessment (Cutting & Scarborough, 2006). As a result, there often is an incomplete understanding of the specific nature of reading problems.

One approach to explaining the various challenges that students face in reading is the Reading Component Model (Aaron, 1997; Gough & Tunmer, 1986). This model provides a way to conceptualize types of struggling readers. This approach separates poor readers into three groups: poor readers with word recognition problems only, poor readers with listening comprehension problems only, and poor readers with a combination of these problems. Readers with primary problems in listening comprehension—that is, the ability to comprehend auditory or spoken information—have been characterized as hyperlexic. These children typically have relatively good word recognition (Gough & Tunmer, 1986).

In a longitudinal study that followed students from kindergarten through fourth grade, Catts, Hogan, and Fey (2003) identified the primary difficulties of children with reading disabilities. They found that approximately 35.5% of children who had reading difficulties had primary difficulty with word recognition, 35.7% had difficulty with both word recognition and listening comprehension, 15.4% had primary difficulty with listening comprehension, and the remaining 13.4% had unspecified difficulty in that they had adequate word recognition and listening comprehension but still performed one standard deviation below the mean on a composite measure of reading comprehension.

Limited word-level skills in decoding can obscure higher-order comprehension skills under traditional methods of large scale assessment. According to the findings of Catts et al. (2003), traditional reading tests might inaccurately measure comprehension, analysis, or interpretation skills of up to about 70% of poor readers who have difficulty with decoding or word recognition. Although older students typically no longer receive instruction on the foundational skills of reading, most reading tests measure growth and proficiency in higher level skills yet still measure proficiency in foundational skills.

Issues in Measurement

Although there are continued debates about the definition of reading proficiency (National Accessible Reading Assessment Projects, 2006), most states' standards include similar sets of reading skills, such as fluency, decoding, and parts of comprehension, that could be measured in annual reading tests. The sets of skills on which state standards tend to focus also change in similar ways across grade levels. As grade levels increase, standards typically focus less on foundational skills such as fluency and decoding and more on higher-order reading skills such as comprehension, analysis, and interpretation. At the higher grade levels, most state reading standards involve recognition of linguistic conventions, thinking or interacting with text, problem solving, and reading as a means of personal growth (Thompson, Johnstone, Thurlow, & Clapper, 2004).

As might be expected, a similar pattern is evident in state reading test specifications and test blueprints. A 2007 survey of standards targeted for measurement by state tests of reading found that 28 states identified specific constructs for their tests (Johnstone, Moen, et al., 2007). The researchers found that by grade 8, many states devoted about 80% of the items in their reading tests to the skills of comprehension, analysis, and interpretation. In contrast, only four states specifically called for measurement of foundational skills such as decoding and word recognition in grade 8, dedicating an average of about 24% of test items to such skills. Foundational reading skills such as decoding are needed to demonstrate proficiency on traditional assessments. If, in the middle and high school grades, the instructional focus has shifted toward primarily higher-order reading skills such as comprehension, one might question the need to continue to include lower-order foundational skills as targets for assessment. Accommodations can remove the need to decode at the same time that students are to show understanding of text.

Accommodations in Reading Assessment

Many students now take tests using a variety of accommodations that are intended to enable them to show their knowledge and skills without the barriers created by their disabilities. An accommodation is a change in materials or procedures that allows the student to show knowledge and skills *without* changing the intent of the measurement (Thurlow, 2007). This general definition has been in the literature for some time (Elliott, Braden, & White, 2001; Koretz & Barton, 2003-2004; Tindal & Fuchs, 2000), with recent clarifications of terminology to distinguish accommodations from assessment modifications which are described as changes in materials or procedures that *do affect* the construct targeted for measurement (Thurlow, Thompson, & Johnstone, 2007).

There is an increasing body of literature concerning the effects of accommodations on test scores. Originally, it was proposed that for an accommodation to yield valid scores, it should result in improved scores for students with disabilities but not for students without disabilities (Phillips, 1994). This notion of an “interaction hypothesis” has been adjusted over time to posit that use of an accommodation may increase test scores of all students, but the benefit is greater for one group (e.g., students with disabilities) more than others (e.g., students without disabilities). This type of interaction effect has been referred to as a “differential boost” (Fuchs & Fuchs, 2001; Laitusis, 2007; Sireci, Scarpati, & Li, 2005).

There has been a fair number of studies conducted on various methods of accommodation in large scale reading assessment, many of which were intended to provide evidence of differential boosts, while others were intended to demonstrate the equivalence of constructs measured. Researchers affiliated with the National Center on Educational Outcomes (NCEO) have completed several comprehensive reviews that examined the research on accommodations (Johnstone, Altman, Thurlow, & Thompson, 2006; Thompson, Blount, & Thurlow, 2002; Zenisky & Sireci, 2007). These reviews covered a variety of works including journal articles, technical reports, and dissertations

that reported the results of studies on accommodations in content areas ranging from math and reading to science or social studies.

Some of the reading accommodations discussed in the NCEO reviews involved alterations in the timing or scheduling of standardized reading tests. In one example, Lesaux, Pearson, and Siegel (2006) studied the effects of timed versus untimed testing conditions, finding that providing extra time increased scores of adults with reading disabilities more than those without reading disabilities on a standardized test of reading comprehension. Similar results were observed by Dicerbo, Stanley, Roberts, and Blanchard (2001) in an experiment on the administration of a standardized reading test over a series of sessions across multiple days. Small yet significant differences were observed between gains experienced by students with disabilities and students without disabilities.

Other accommodations for reading assessment alter the presentation of a test. One example involves the presentation of a reading test in a computerized rather than a standard paper and pencil format. Pomplun, Frey, and Becker (2002) found some small evidence to suggest that presenting a reading test on a computer could produce higher scores in vocabulary and comprehension for high school and college students without disabilities. In a study on the effect of computerized presentation on test taking strategies, Kobrin and Young (2003) observed similar scores in either testing condition.

Another commonly studied accommodation for reading assessments is the “read aloud” accommodation. This accommodation involves the presentation of entire tests in an auditory format, including passages and test items, or just the test items (questions and response choices), in addition to the printed text. The intention is to remove decoding demands for test takers, thus allowing students who struggle with foundational reading skills to demonstrate their higher-order reading skills.

Research on this accommodation has yielded mixed results. As reviewed by Hollenbeck (2002), some studies yielded evidence that the read aloud accommodation produces significant gains on math tests for students with disabilities or students with low reading skills (e.g., Helwig, Rozek-Tedesco, Tindal, Heath, & Almond, 1999; Tindal, Heath, Hollenbeck, Almond, & Harniss, 1998; Weston, 1999). Other studies yielded no significant gains (Hollenbeck, Rozek-Tedesco, Tindal, & Glasgow, 2000; Tindal, Almond, Heath, & Tedesco, 1998; Tindal, Glasgow, Helwig, Hollenbeck, & Heath, 1998). Laitusis (in press) administered reading comprehension tests under standard and read aloud conditions to students with and without reading disabilities in grades 4 and 8. Both groups of students had higher mean scores with the audio presentation, but students with reading disabilities benefited more than students with no disabilities, thereby demonstrating the differential boost effect.

Findings from research on the equivalence of scores from tests read out loud to tests presented in the traditional manner have been equally diverse. Due to the potential complexity of interaction between characteristics of tests, accommodations, and

students, it has been difficult for researchers to determine how accommodations such as the “read aloud” affect the validity and accuracy of student scores. Huynh and Barton used error analyses (Barton & Huynh, 2003) and factor analyses (Huynh & Barton, 2006) to conclude that use of a read aloud accommodation does not alter the tested construct. On the other hand, results from research analyzing differential item functioning suggest that read aloud accommodations may be inappropriate for reading assessments because the accommodations change the construct being measured (Bolt & Ysseldyke, 2006).

Given the mixed results, presenting the test out loud to the student continues to be one of the more controversial methods to accommodate reading tests. This is evident in complex and varying state policies for these accommodations (Christensen, Lazarus, Crone, & Thurlow, 2008; Lazarus, Thurlow, Lail, & Christensen, 2009; Thurlow, 2007). Some policies indicate that the read aloud accommodation alters the construct being measured (for example, from reading comprehension to listening comprehension) or that the accommodation artificially increases the scores of all students, regardless of disability, while others indicate that the read aloud accommodation does not change the construct of comprehension and therefore is allowable on all items except those measuring decoding skills.

Partial Auditory Accommodation

Most of the literature on auditory accommodations has used an approach in which the entire printed text is read aloud. This might be referred to as a “full auditory accommodation” abbreviated as “FAA.” A different kind of accommodation that might be labeled a “partial auditory accommodation” or “PAA” is one in which isolated words or phrases are pronounced aloud, most likely in response to the reader’s selection. We see several important distinctions between FAA and PAA approaches. First, by working with one word at a time, PAA approaches are focused on supplementing only those skills that are needed to derive word level pronunciation. This contrasts with FAA approaches, which can introduce much greater levels of support for comprehension through the vocal cues that are part of effective prosody. A second point is that having the student choose which words or phrases are read, if any, gives the student control of the accommodation. Our hypothesis was that giving students more control over the auditory accommodation would maximize the goodness of fit between the accommodation and individual student need. Part of this control is that students determine how they access the text. With PAA, students can easily return to specific words or specific sections of a passage. A related hypothesis was that PAA more closely aligns with reading under typical conditions and consequently does less to undermine affective domains such as motivation and self efficacy.

For those struggling readers who have skills in decoding, word identification, and fluency sufficient to access most of the test content and items, having a small portion of the text pronounced aloud should be less of a threat to validity, and should also help students access meaning in test content and items. Testing in this format would still require students to demonstrate proficiency in the higher-order reading skills of

comprehension, analysis, and interpretation.

Read-aloud accommodations have commonly relied on the use of tape recorders, compact disks, or human readers to present tests to students (Christensen et al., 2008). These approaches almost always constitute full auditory accommodations because they are not readily adjusted to be a partial auditory presentation. There is technology available, however, that would allow for a PAA. An example of such technology, usable with paper and pencil tests, is a reading pen—a device that uses optical character recognition and speech synthesis capabilities to facilitate reading comprehension.

Preliminary evidence suggests that reading pens can be beneficial to comprehension for students with learning disabilities. Higgins and Raskind (2005) trained 34 students with disabilities between the ages of 10 and 18 to use a reading pen. During a two week familiarization period, the researchers collected observational data on student use of the pen's features. They reported that students accessed definition and syllabification features about 20% of the time students were observed using the pen to scan words. After the familiarization period researchers administered a text-based, standardized test of reading comprehension to each student twice—once with and once without the reading pen. Results indicated significant and moderate gains in performance when students were able to use the reading pen.

Although Higgins and Raskind (2005) indicated that they were examining the effectiveness of the reading pen as a way to address students' word-level skill deficits (decoding and word identification), students were allowed to use the pen to obtain definitions and synonyms during testing. The definition and synonym function addresses the text-based skills of vocabulary and comprehension. No data were reported that described the frequency with which students used the pen for these purposes, and it is possible that the results represented compensation not only for word-level skill deficits, but also for deficits in higher-order reading skills. Also, because only students with disabilities were included in the sample, there was no evidence that this type of technology can be used to accommodate students with specific types of skill deficits.

The goal of the present study was to evaluate the effectiveness of a reading pen as a PAA for large scale reading assessments when the pen is used only to pronounce words on demand. The vocabulary function on the pen was not employed during this study. In addition, the reading pen was used by both students with disabilities and students without disabilities to check for a differential boost. We hypothesized, further, that if the reading pen was effective at compensating for deficits in decoding and fluency skills, it would be less effective for stronger readers and more effective for students with disabilities related to reading. Four research questions were addressed:

1. To what extent does use of the reading pen on a standardized test of reading affect scores?

2. To what extent is there a different effect for students with disabilities and students without disabilities?
3. Regardless of disability status, to what extent does use of the reading pen affect the scores of students with adequate fluency?
4. What are students' perceptions of the helpfulness of the reading pen?

Methods

Participants

Students in grades 6 and 8 from two schools within a large suburban district in the Midwest served as participants in this study. Students receiving special education services (referred to as students with disabilities) and students not receiving special education services (referred to as students without disabilities) were both included; 44 students without disabilities were recruited from general education reading classes, and 32 students with disabilities were recruited from special education reading classes. Although information regarding students' specific diagnoses was not collected, all students with disabilities in this sample were receiving special educational services in reading. The total sample of 76 students included 46 students in grade 6 and 30 students in grade 8.

Student demographic information was provided by the school. Of the total sample, 1 student was Asian, 15 were Black, 8 were Hispanic, and 47 were White. The demographic information for five students was unavailable. After district and school-level permissions were obtained, students were recruited via informational flyers passed out by teachers during class. Students who returned permission slips signed by a parent or guardian were invited to participate.

Measures

Estimates of oral reading fluency were obtained to help determine the extent to which use of the reading pen interacted with deficits in decoding and word recognition. Standard, graded AIMSweb curriculum-based measurement reading probes (CBM-R) were administered. Shinn and Shinn (2002) reported that the AIMSweb passages have adequate technical features necessary for making decisions about students' general reading skills. In addition to measuring fluency, three AIMSweb maze reading tasks were administered to obtain brief measures of broad reading skills including comprehension. Research supports the use of maze tasks as a reliable, sensitive, and valid assessment procedure of reading (Shinn, Deno, & Espin, 2000). All AIMSWeb materials were used with permission of the publisher.

Student reading comprehension in test conditions with and without the reading pen was measured with a commercially available standardized reading test. The Gray Silent

Reading Test (GSRT) is a norm-referenced assessment of reading comprehension with estimates of internal consistency above .90 and average test-retest reliability of .85 (Wiederholt & Blalock, 2000). The GSRT consists of a series of short passages ordered by difficulty level (from least to most challenging), which enables raw scores to be used to calculate a Reading Quotient. GSRT Reading Quotients are standard scores with a mean of 100 and a standard deviation of 15. The standard errors of measurement reported for GSRT Reading Quotients range between three and four standard scores.

Although this test was designed to be individually administered, the test's authors provided guidance on appropriate methods of group administration; due to time constraints, group administrations of the GSRT were used for this study. Students were administered forms A and B of the GSRT in a randomized order. Students read a predetermined set of stories covering what was deemed to be an appropriate range of difficulties for the grades within the sample and answered all items. Standard scoring rules were applied to determine Reading Quotients. Testing with the GSRT took place in a single session; total administration time under group administration required about 25 minutes for each form, totaling approximately 50 minutes.

Materials

Materials in this study included reading pens. The Reading Pen II can be used by students to scan printed text and receive audible pronunciations, definitions, and synonyms of the scanned words. The pen is a battery powered hand-held device, adjustable for left and right handed users, that pairs infrared optical character recognition with speech synthesis capabilities, and is capable of recognizing and pronouncing nearly 250,000 words in both English and Spanish. Each user was issued three components: a reading pen, a plastic guide intended to provide assistance in scanning text, and a pair of ear buds to hear text read out loud. Users scan text by rolling the pen over a single word or a series of words they intend to have read out loud. After the optical sensor recognizes a specific word it is displayed in a small LCD screen, and pronounced audibly through ear buds.

Procedures

Students were first asked to read three CBM-R probes. Using procedures recommended by Shinn and Shinn (2002), each student was individually administered three grade level passages for one minute each. As a student read, the researcher marked words read incorrectly. Scores obtained were words read correctly per minute (WRC). The median of three CBM-R scores was recorded. Second, students completed three CBM maze tasks. Maze passages have brackets containing three bolded word options every seventh word. One of the words is correct and the other two are distractors. Students had 3 minutes to read and choose words in the brackets (indicated by circling or highlighting the word). Scores used in analysis were the number of correct answers and the number of incorrect answers.

The third activity was training on the reading pen. Researchers followed a standardized

training procedure that included an instructions handout and a worksheet on which students practiced using the pen (see Appendices A and B). For approximately 15 minutes, researchers guided students through instructions and practice with the pen. Researchers observed students while they practiced the pen and provided help for any students who were having difficulties. Prior to moving to the testing stage, students were asked whether they felt comfortable using the pen on the test. If any student indicated that he or she did not feel comfortable, researchers spent more time helping the student practice until the student was proficient in using the pen. Typically, students learned to use the pen in 5-10 minutes.

The fourth activity was the Gray Silent Reading Test (GSRT). Each student took both form A and B, one form with the pen and one form without, with counterbalancing of forms and order of condition (with pen or without pen) using a random number generator in the Microsoft Windows Excel spreadsheet program. Students had 25 minutes to complete passages 4-11 of each form. These passages were chosen based on a recommendation within the test manual for methods of group administration (p. 17).

The last activity was a short questionnaire about the reading pen (see Appendix C) intended to measure the extent to which students liked the device. Students rated the pen's helpfulness on a scale ranging from A (very helpful) to D (not helpful); responses were scored according to the degree of helpfulness (A = 4, D = 1). The questionnaire included a second open-ended question to solicit students' thoughts on what aspects or features of the device were helpful or needed improvements. Once the questionnaires were returned to researchers, students were given gift cards to a local store in appreciation for their participation in the study.

Results

Effects of the Reading Pen

Table 1 shows the results of testing for each group (no disability, disability), and for each condition (no pen, reading pen). Students without disabilities performed higher than students with disabilities on the GSRT regardless of condition. A repeated measures analysis of variance indicated a significant between-group effect but did not indicate any significant within-group differences. A dependent samples t-test that compared only the performance of the group of students with disabilities resulted in a small significant difference between testing conditions, $t(41) = -2.55, p < .05$, Cohen's $d = -0.33$.

Although there were no main effects observed for use of the pen for the entire group of students in this study, potential positive effects for a smaller subset of students cannot be ruled out. Table 2 presents information about students who might have benefited from using the pen compared to those whose test scores did not appear to differ substantially. The GSRT Examiner's Manual specifies that two GSRT Reading Quotients should only be considered significantly different at a .05 level when a difference of nine or more

Table 1. GSRT Scores by Test Condition

Student Group	GSRT Reading Quotients Mean (SD)		Student Group
	No Reading Pen	Reading Pen	Mean (SD)
No Disability N = 32	100.81 (17.83)	101.34 (17.58)	101.08 (17.56)
Disability N = 44	80.55 (11.42)	76.89 (10.66)	78.72 (11.14)
Condition Mean (SD)	89.08 (17.54)	87.18 (18.46)	

Table 2. Demographic Characteristics of Students who Improved^a with the Reading Pen and Those Who Did Not

	Grade 6	Grade 8	Male	Female	Disability	No Disability
Improved with Pen (n = 11)	7	4	6	5	5	6
Did Not Improve with Pen (n = 65)	69	72	70	71	71	70

^aImprovement was defined as scoring 9 or more points higher on the test with the pen than on the test without.

points is observed (Weiderholt & Blalock, 2000). The difference between the GSRT score with and without the reading pen was calculated for each student. For the majority of students, scores obtained with the reading pen were less than nine points greater than those obtained under the standard presentation. Despite this, 11 students did obtain scores above the nine point criterion while using the reading pen. We were not able to identify any distinguishing characteristics of those 11 students by looking at data on grade, gender, or special education status.

CBM-R and Maze

Because Maze scores were highly correlated with oral reading fluency, as measured by CBM-R, only the CBM-R results are reported here (see Table 3). No significant differences in oral reading fluency were found by grade, but a significant difference in reading fluency was observed between students by disability status, $t(71) = 8.724$, $p < .001$; the performance of students with disabilities was lower than that of students without disabilities. As shown in Table 3, the standard deviations for students with disabilities were slightly smaller than for students without disabilities.

Additional analyses indicated that the CBM-R performance was correlated at a moderate

Table 3. CBM-R Measures of Oral Reading Fluency by Disability Status and Grade

Student Group	Grade	
	6 M (SD)	8 M (SD)
No Disability	162.90 (38.12)	143.27 (35.87)
Disability	88.25 (33.19)	83.44 (26.73)

level with performance on the GSRT, regardless of whether the reading pen was used ($r = .63$) or not used ($r = .65$). This result is reflected in Table 4, which shows the performance of all students with and without the pen when students were grouped according to their CBM-R percentile rank.

Student Evaluations of the Reading Pen

Students' responses to a questionnaire about the reading pen were reviewed to determine their perspectives on the effectiveness of this accommodation. The survey contained two questions. Students responded to the first question ("Was the reading pen helpful for you on the test?") using a Likert-type scale (A = very helpful, B = pretty helpful, C = kind of helpful, and D = not helpful). These letter responses were coded as numbers: A was 4, D was 1. The second question asked students to provide any other feedback about the reading pen.

Of the 76 participants, 72 questionnaire responses were obtained, but only 52 of these responses could be linked to individual participants because of a lack of identification information recorded on some of the surveys that students completed. The students whose questionnaire responses could not be linked included 4 with a disability, 20 with no disability, 13 in grade 6, and 11 in grade 8. We present the overall descriptive data, and then the data for only those students for whom it was possible to link scores to responses.

Students' ratings of the helpfulness of the reading pen were generally more positive than

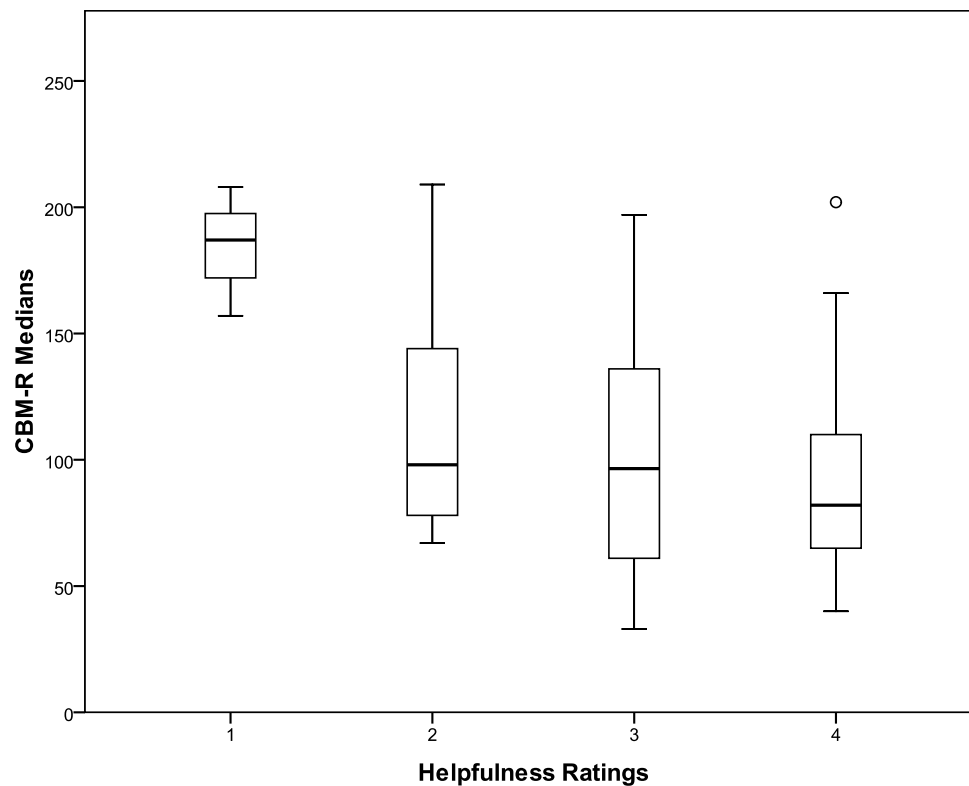
Table 4. Average GSRT Reading Quotients by Oral Reading Fluency (ORF) Percentile Rank

GSRT Means (SD) by Testing Condition		
ORF Percentile Group	Reading Pen	No Reading Pen
1% - 20%	75.50 (15.74)	76.36 (17.03)
20% - 40%	78.94 (8.93)	81.31 (5.57)
40% - 60%	81.00 (13.77)	83.36 (9.84)
60% - 80%	92.40 (18.92)	98.13 (14.35)
80% - 90%	105.06 (16.37)	103.59 (19.69)

negative ($M = 2.88$, $SD = 0.87$). Out of the 72 responses, 19 students rated the pen as being “very helpful,” 29 rated it as “pretty helpful,” 20 rated it as “kind of helpful,” and 4 students rated the pen as “not helpful.” Analysis of the relationship between helpfulness ratings and changes in GSRT scores from the no pen to the pen condition showed no relationship ($r = 0.19$, $n = 51$).

Figure 1 shows the relationship between ratings of the pen’s helpfulness and oral reading fluency for those students whose questionnaire responses could be linked to testing data. There was a significant negative correlation between helpfulness rating and oral reading fluency ($r = -0.344$, $n = 51$). Students with higher fluency tended to rate the pen as being less helpful than students with lower oral reading fluency. This trend did not appear to differ substantially according to educational placement or gender.

Figure 1. Relationship Between Students’ CBM-R Medians and Their Ratings of the Pen’s Helpfulness



Students’ open-ended responses addressed which aspects of the pen they found to be helpful and ideas for improving the pen. Twelve student responses merely indicated correct use of pen, such as scanning words they did not know or could not pronounce. Of three students who suggested that the pen would help more if it provided word definitions, one was from the group whose GSRT score improved. None of 12 students who made suggestions for improving technical features of the pen, such as audio quality

or scanning precision, received a higher score when using the pen. Similarly, none of six students who commented that the reading pen would likely help students who are poor readers but not students who are good readers obtained higher scores when using the reading pen.

Discussion

This study was conducted to explore the potential usefulness of a partial auditory accommodation (PAA) for students who have disabilities related to reading during a standardized test of reading comprehension. There have been increasing concerns about the assessment of students for whom traditional tests of reading comprehension are less accessible (or accurate) due to barriers imposed by deficits in word level skills such as decoding. This issue is commonly addressed through the use of full auditory accommodations (FAA), although research on this type of accommodation has produced mixed results. A sample of 76 students from grades 6 and 8 were administered two forms of a standardized test of reading comprehension—one with and one without use of a reading pen as a partial auditory accommodation.

Our results suggest that a PAA, at least in this study, was not useful for students with disabilities that affect reading. This statement does not apply to every student in the sample because there were some students who appeared to benefit from the accommodation. We find it difficult to make a case for considering these results for students who appeared to benefit from using the reading pen to be more than statistical noise because we have not been able to identify characteristics in the data we had available for this sample of students that distinguish students who benefited from those who did not. Our overall results are different from results obtained by Higgins and Raskind (2005), who reported a moderate effect size (Cohen's $d = .69$) for students with disabilities using a reading pen. Differences in how the reading pen was used in the two studies might explain the differences in findings. In the study by Higgins and Raskind, students were able to use the reading pen to obtain word definitions and synonyms, which would compensate not only for deficits in decoding and word recognition, but also for vocabulary. In the current study, the definition and synonym functions were blocked so students only were able to hear the word or phrase read out loud, limiting the focus of the accommodation to word-level deficits in decoding. It seems plausible that the effect observed by Higgins and Raskind (2005) was due at least in part to assistance with vocabulary in addition to decoding.

Although our study did not yield significantly different test scores resulting from use of the reading pen, it is possible that a subset of students might benefit from its use. Researchers have noted that there is heterogeneity among students with disabilities (Salvia & Ysseldyke, 2006; Ysseldyke & Algozzine, 2006), and that practices designed for students with identical diagnoses will not have uniform effects (Fuchs & Fuchs, 1999; Morris et al., 1998). It is possible that improved results from a subset of students with disabilities were obscured by the scores of other students who did not benefit from

use of the pen. Indeed, results indicated that 11 students did obtain scores 9 or more points higher with the reading pen than without; 5 of these students were students with disabilities receiving special education services.

It also is possible that effects of the reading pen may have been diminished by certain elements of the study. The length of time students were given to familiarize themselves with the technology might not have been enough. Higgins and Raskind (2005) provided students about two weeks for familiarization and practice with the reading pen. Students in the current study demonstrated competent use of the pen within the time provided for practice (approximately 15 minutes), but it is possible that the additional practice provided by Higgins and Raskind would have been beneficial for some students in our study.

The process for selecting students with disabilities for our study also might have limited the likelihood of recruiting students for whom the reading pen could be useful. Available literature indicates that up to about one third of students known to have skill deficits in reading experience difficulties as a result of limited decoding or word recognition (Catts et al., 2003). Use of the CBM-R measure, which required students to read passages, may have underestimated the decoding problems of some students because the passage could have provided some contextual cues to assist students with decoding. It may be that the use of nonsense words or word recognition tasks would have better revealed decoding problems that would, in turn, relate to the performance of students with the reading pen.

A second issue to consider is the effect of fluency on comprehension. Researchers have found evidence that comprehension of text at an elementary difficulty level requires levels of fluency ranging from 45 to 99 words per minute (Burns et al., 2002). We found that oral reading rates of students without disabilities were much higher than those of students with disabilities, in both grades (163 vs. 88 in grade 6 and 143 vs. 83 in grade 8). With average rates below 90 words per minute, it is possible that boosts from use of the reading pen for students with disabilities were masked by fluency limitations. The aim of the study was to remove barriers due to decoding and word recognition limitations, not necessarily fluency limitations. Furthermore, it may be that using the pen slowed some students down. Use of the pen could have helped some students overcome difficulties with decoding unfamiliar words, but increases in reading accuracy did not offset deficits in reading rate. It appeared that some students had difficulty finishing the tests on time, suggesting that low fluency also could have limited performance by preventing students from spending enough time on each item or rushing at the end.

Conclusion

Based on the results of this study, it cannot be concluded that use of a reading pen as a PAA affects students' test scores either positively or negatively. Our hypothesis was that by allowing students to use the reading pen to pronounce words that were unfamiliar or difficult to decode, students with disabilities related to reading would have a greater opportunity to apply their higher-order skills in comprehension, and could therefore be

assessed more accurately. Although no group differences due to use of the reading pen were observed, the possibility that a number of students could have benefitted from its use cannot be ruled out.

Similarly, a number of students could have had deficits in reading skills in addition to decoding. For instance, it is possible that deficits in vocabulary limited the extent to which students could benefit from having single words pronounced automatically. Had students been allowed to access definitions of unfamiliar words, testing results might have been more similar to those obtained by Higgins and Raskind (2005). Yet had the pen been used to compensate for vocabulary in addition to potential deficits in decoding and word recognition, its effectiveness as an accommodation would be diminished.

Limitations and Future Research

Several limitations potentially affected the results of this study. One limitation was the sample of participants, which came from two schools in the same district. Students with disabilities had comparable numbers of students in each of the two grades from the two schools. In contrast, students without disabilities were not equally distributed across grades across the two schools, with one school providing the grade 6 students and the other school providing the grade 8 students. This sampling pattern confounds the results with possible school-level effects.

Another limitation in this study was the nature of the data collected on word-level reading skills. Measures of oral reading fluency require both decoding of unfamiliar words and automatic recognition of familiar words. It is possible that some participants were able to recognize familiar words in the reading fluency probes that they might not have been able to decode correctly had the words been unfamiliar. This would obscure the extent to which use of the pen interacts with word-level reading skills measured by oral reading fluency probes. Additional research on partial auditory accommodations should consider using more specific measures of word-level reading subskills such as nonsense word fluency, a more direct measurement of a student's ability to decode text.

Another limitation of this study was the extent to which students were able to demonstrate competent use of the pen. Researchers led participants through a brief activity at the beginning of the study. Each student was encouraged to ask questions if he or she encountered any difficulties with the pen, and each was asked to demonstrate that he or she could successfully scan a word and hear the word pronounced through the ear buds. Still, this did not illustrate the extent to which students incorporated the pen into their existing strategies for reading comprehension. Higgins and Raskind (2005) provided students with two weeks of practice, whereas the exposure and practice period provided in this study lasted for about 15 minutes. It is possible that amount of time might not have been sufficient for students to become familiar with the functions and utility of the reading pen. Future studies on use of novel technology in test accommodations should include methods to verify that students have learned to use the accommodation as intended.

References

- Aaron, P. G. (1997). The impending demise of the discrepancy formula. *Review of Educational Research*, 67, 461-502.
- Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of student language background on content-based performance: Analyses of extant data* (CSE Technical Report No. 603). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Barton, K., & Huynh, H. (2003). Patterns of errors made by students with disabilities on a reading test with oral reading administration. *Educational and Psychological Measurement*, 63, 602-614.
- Bolt, S., & Ysseldyke, J. (2006). Comparing DIF across math and reading/language arts tests for students receiving a read-aloud accommodation. *Applied Measurement in Education*, 19, 329-355.
- Burns, M. K., Tucker, J. A., Hauser, A., Thelen, R. L., Holmes, K. J., & White, K. (2002). Minimum reading fluency rate necessary for comprehension: a potential criterion for curriculum-based assessments. *Assessment for Effective Intervention*, 28, 1-7.
- Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities*, 36, 151-164.
- Christensen, L. L., Lazarus, S. S., Crone, M., & Thurlow, M. L. (2008). *2007 state policies on assessment participation and accommodations for students with disabilities* (Synthesis Report 69). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10(3), 277-299.
- Darling-Hammond, L. (2003). Standards and assessments: Where we are and what we need. *Teachers College Record*. Retrieved April 20, 2009 from <http://www.tcrecord.org/Content.asp?ContentID=11109>
- Dicerbo, K. E., Stanley, E., Roberts, M., & Blanchard, J. (2001). *Attention and standardized reading test performance: Implications for accommodation*. Paper session presented at the annual meeting of National Association of School Psychologists, Washington, DC.

- Dolan, R.P., Hall, T.E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *Journal of Technology, Learning, and Assessment*, 3(7).
- Elbaum, B., Arguelles, M. E., Campbell, Y., & Saleh, M. B. (2004). Effects of a student-reads-aloud accommodation on the performance of students with and without learning disabilities on a test of reading comprehension. *Exceptionality*, 12, 71-87.
- Elliott, S. N., Braden, J. P., & White, J. L. (2001). *Assessing one and all: Educational accountability for students with disabilities*. Alexandria, VA: Council for Exceptional Children.
- Fuchs, L. S., & Fuchs, D. (1999). Fair and unfair testing accommodations. *School Administrator*, 56, 24-29.
- 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.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6-10.
- Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. (1999). Reading as an access to mathematics problem solving on multiple choice tests for sixth-grade students. *Journal of Educational Research*, 93, 112-125.
- Higgins, E. L., & Raskind, M. H. (2005). The compensatory effectiveness of the Quictionary Reading Pen II on the reading comprehension of students with learning disabilities. *Journal of Special Education Technology*, 20(1), 31-40.
- Hollenbeck, K. (2002). Determining when test alterations are valid accommodations or modifications for large-scale assessment. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students* (pp. 395-425). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hollenbeck, K., Rozek-Tedesco, M., Tindal, G., & Glasgow, A. (2000). *Computation as a predictor of large-scale tests scores*. Unpublished manuscript, University of Oregon.
- Huynh, H., & Barton, K. (2006). Performance of students with disabilities under regular and oral administrations of a high-stakes reading examination. *Applied Measurement in Education*, 19(1), 21-39.
- Individuals with Disabilities Act. (2004). Public Law 108-446. Washington, DC: U. S. Government Printing Office.

- Johnstone, C. J. (2003). *Improving the validity of large-scale tests: Universal design and student performance* (Technical Report 37). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Johnstone, C. J., Altman, J., Thurlow, M. L., & Thompson, S. J. (2006). *A summary of research on the effects of test accommodations: 2002 through 2004* (Technical Report 45). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Johnstone, C. J., Moen, R. E., Thurlow, M. L., Matchett, D., Hausmann, K. E., & Scullin, S. (2007). *What do state reading test specifications specify?* Minneapolis, MN: University of Minnesota, Partnership for Accessible Reading Assessment.
- Kobrin, J. L., & Young, J. W. (2003). The cognitive equivalence of reading comprehension test items via computerized and paper-and-pencil administration. *Applied Measurement in Education*, 16(2), 115-140.
- Koretz, D., & Barton, K. (2003-2004). Assessing students with disabilities: issues and evidence. *Educational Assessment*, 9, 29-60.
- Laitusis, C. C. (2007). Research designs and analysis for studying accommodations on assessments. In C. C. Laitusis & L. L. Cook (Eds.), *Large-scale assessment and accommodations: What works?* Arlington, VA: Council for Exceptional Children.
- Laitusis, C. C. (in press). Examining the impact of audio presentation on tests of reading comprehension. *Applied Measurement in Education*.
- Lazarus, S. S., Thurlow, M. L., Lail, K. E., & Christenson, L. (2009). A longitudinal analysis of state accommodations policies: Twelve years of change, 1993-2005. *Journal of Special Education*. 43(2), 67-80.
- Lesaux, N. K., Pearson, M. R., & Siegel, L. S. (2006). The effects of timed and untimed testing conditions on the reading comprehension performance of adults with learning disabilities. *Reading and Writing*, 19, 21-48.
- Moen, R., Liu, K., Thurlow, M., Lekwa, A., Scullin, S. & Hausmann, K. (2009). Identifying less accurately measured students. *Journal of Applied Testing Technology*, 10(2).
- Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., et al., (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology*, 90, 347-373.
- National Assessment Governing Board (2008). *Reading framework for the 2009 National Assessment of Educational Progress*. Washington, DC: U.S. Government Printing Office. See <http://www.nagb.org/publications/frameworks/reading09.pdf>

National Accessible Reading Assessment Projects. (2006). *Defining reading proficiency for accessible large scale assessments: Some guiding principles and issues*. Minneapolis, MN: Author.

National Reading Panel. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

No Child Left Behind Act. (2001). Public Law 107-110. Washington, DC: U. S. Government Printing Office.

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

Pomplun, M., Frey, S., & Becker, D. F. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement*, 62(2), 337-354.

Salvia, J., & Ysseldyke, J.E. (2006). *Assessment in special education and inclusive education*. Boston: Houghton-Mifflin.

Shin, J., Deno, S. L., & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *The Journal of Special Education*, 34, 164-172.

Shinn, M. R., & Shinn, M. M. (2002). *Administration and scoring of Reading Curriculum Based Measurement (R-CBM) for use in general outcome measurement*. Eden Prairie, MN: Edformation Inc.

Shriner, J. G., Spande, G., & Thurlow, M. L. (1994). *State special education outcomes 1993*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Sireci, S. G., Scarpata, 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.

Stahl, S. A., & Hiebert, E. H. (2005). The word factors: A problem for reading comprehension assessment. In S. G. Pearson & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 161-186). Mahwah, NJ: Lawrence Erlbaum Associates.

Thompson, S., Blount, A., & Thurlow, M. (2002). *A summary of research on the effects of test accommodations: 1999 through 2001* (Technical Report 34). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thompson, S., Johnstone, C., & Thurlow, M. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thompson, S. J., Johnstone, C. J., Thurlow, M. L., & Clapper, A. T. (2004). *State literacy standards, practice, and testing: Exploring accessibility* (Technical Report 38). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thompson, S. J., Thurlow, M. L., & Malouf, D. (2004). Creating better tests for everyone through universally designed assessments. *Journal of Applied Testing Technology*, 10(2). See <http://www.testpublishers.org/atp.journal.htm>

Thurlow, M. L. (in press). Steps toward creating fully accessible reading assessments. *Applied Measurement in Education*.

Thurlow, M. L. (2007). State policies and accommodations: Issues and implications. In C. Cahalan-Laitusis & L. Cook (Eds.), *Accommodating students with disabilities on state assessments: What works?* Arlington, VA: Council for Exceptional Children.

Thurlow, M., Bremer, C., & Albus, D. (2008). *Good news and bad news in disaggregated subgroup reporting to the public on 2005-2006 assessment results* (Technical Report 52). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L., Laitusis, C. C., Dillon, D. R., Cook, L. L., Moen, R. E., Abedi, J., & O'Brien, D. G. (2009). *Accessibility principles for reading assessments*. Minneapolis, MN: National Accessible Reading Assessment Projects.

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: University of Minnesota, Partnership for Accessible Reading Assessment.

Thurlow, M. L., Moen, R., & Wiley, H. I. (2005). *Annual performance reports: 2002-2003 state assessment data*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. See <http://www.nceo.info/OnlinePubs/APRsummary2005.pdf>

Thurlow, M. L., Quenemoen, R., Altman, J., & Cuthbert, M. (2008). *Trends in the participation and performance of students with disabilities* (Technical Report 50). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M., Thompson, S., & Johnstone, C. (2007). Policy, legal, and implementation issues surrounding assessment accommodations for students with disabilities. In L. Florian (Ed.), *The Sage handbook of special education*. Thousand Oaks, CA: Sage.

Tindal, G., Almond, P., Heath, B., & Tedesco, M. (1998). *Single subject research using audio cassette read-aloud in math*. Unpublished manuscript, University of Oregon.

Tindal, G., & Fuchs, L. (2000). *A Summary of research on test changes: An empirical basis for defining accommodations*. Lexington, KY: University of Kentucky, Mid-South Regional Resource Center. (ERIC Document Reproduction Service No. ED 442 245)

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

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

Weston, T. (1999, April). *The validity of oral presentation in testing*. Paper presented at the annual conference of the American Educational Research Association, Montreal, Canada.

Wiederholt, J. L., & Blalock, G. (2000). *Gray Silent Reading Tests: Examiner's Manual*. Austin, Texas: Pro-ed.

Ysseldyke, J. E., & Algozzine, B. (2006). *Public policy, school reform, and special education: A practical guide for every teacher*. Thousand Oaks, CA: Corwin Press.

Zenisky, A. L., & Sireci, S. G. (2007). *A summary of the research on the effects of test accommodations: 2005-2006* (Technical Report 47). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Appendix A

The Reading Pen Instruction Packet

A. The Reading Pen Can:

- Scan words by rolling the pen over them
- Say words out loud

B. Parts

Display
Keypad
Rollar Tip
Ear Bud

Buttons

Power
Enter
Escape
Arrows

Turn on

Press the Power button

C. How to Scan:

1. Is the red light on the tip of the pen blinking?
If yes, you are ready to go
If no, press the Power Button
2. The **home screen** says “Scan in English” on the top.
3. You should always keep the words in the trainer’s “mouth”.
4. When you scan, include the spaces before and after the word.
5. **You do not need to press down on the pen.** Make sure that the word on the pen screen matches the word that you scanned exactly. If it doesn’t, simply scan the word again. You do not have to push any buttons to scan again.

D. How to hear the word:

1. First you scan a word. It will say the word out loud after it scans it.
2. Press the ENT button to hear the word repeated.

E. How to scan a phrase or sentence:

1. Scan the phrase or sentence just like you would for a word.

2. The pen will say the entire phrase or sentence.
3. If you press the ENT button after you scan, it will only repeat the first word.
4. If you want to hear individual words, use the arrow buttons to move to the word.
5. To hear the entire sentence read out loud together, you will have to rescan.

G. Using the Pen on the Test:

1. Do not use the pen in any other way than what you have just been taught. You do not have to use the pen if you don't want to.
2. If at any time you are stuck and can't get back to the Home Screen, just turn off your pen and then turn it on again.
3. We cannot answer any questions about the test, but if you have any questions about the pen please raise your hand and we will come over to help you.
4. After the test, you will have time to play with the pen if you would like. Please use the time we give you for the test only for the test – do not play with the pen until after you are done.

Appendix B

Practice Paper

1. When you scan, at the space **before** the word and go to the space **beyond** the word. Start and stop at the lines below. Now practice scanning this word:

[**Practice**]

Does the word "Practice" show up on the screen? If it doesn't, try again! You do not have to clear the screen or press any buttons before trying again. Always make sure that the word on the screen matches the word on the paper exactly.

2. Keep practicing on the words below. Make sure that the word comes up on the screen. Raise your hands if you have any questions.

**Reading
School**

Now practice scanning the word when it is in a sentence.

Scan the word "Teacher".

The teacher was happy because her students worked hard.

Scan the word "Happy".

The teacher was happy because her students worked hard.

Put your pen down when you have finished all of the words.

3. Scan the word below. Then press the ENT button to hear the word again.

Weather

Try some more. Raise your hand if you have questions. When you are done, put your pen down.

Scan the word "**succeed**".

The students will succeed because they are dedicated to school.

Scan the word "**dedicated**".

The students will succeed because they are dedicated to school.

Please put your pen down when you are done.

Appendix C

1. Was the reading pen helpful for you on the test?

- A. Very helpful
- B. Pretty helpful
- C. Kind of helpful
- D. Not helpful

2. In the box below, please write down any comments or feedback you have for us about the reading pen.

Thank you for helping us learn how to make reading tests better for students!