

Use of Audio Modification in Science Vocabulary Assessment

Tufan Adiguzel

Bahçeşehir Üniversitesi, İstanbul, TURKEY

Received 13 January 2011; accepted 27 July 2011

The purposes of this study were to examine the utilization of audio modification in vocabulary assessment in school subject areas, specifically in elementary science, and to present a web-based key vocabulary assessment tool for the elementary school level. Audio-recorded readings were used to replace independent student readings as the task demand for progress monitoring science vocabulary for 162 fifth grade students in 14 schools. Scores on vocabulary tests obtained by students who used the audio texts for readings were compared with scores obtained by the same students who did not use the audio texts. A statistically significant difference between the two scores in favor of the students who used the audio texts was found. Benefits for at-risk students of school failure, and the trend between the amount of audio use and test scores were also investigated. Results suggest that, if appropriately conducted, the use of this type of modification together with written text may enable students feel confident about their science vocabulary knowledge.

Keywords: Read Aloud, Test Accommodation, Key Vocabulary, Science Assessment, Internet

INTRODUCTION

Given the prominence of vocabulary in learning to read, vocabulary assessment can be seen as a very important component and continuous process in specific content area lessons, especially at the elementary level. *How much* and *how well* the students achieve in vocabulary are crucial (Braun, 2009; Pearson, Hiebert, and Kamil, 2007). In fact, in their research, Pearson, Hiebert, and Kamil (2007) argue vehemently that vocabulary assessment is clearly being given little attention—both in its theoretical and experiential forms. Pearson, Hiebert, and Kamil (2007) also argue that the study of vocabulary has only been driven by tradition, ease, psychometrics, and a mere search for economy of effort instead of a rich conceptualization of its nature and connection to other facets of reading proficiency,

and more importantly comprehension. Therefore, explicit studies on vocabulary assessment, in its entire form and with regard to its relevance to learning, are needed. However, only a few, but obviously meager, studies have been conducted.

Over the past decade, a concern about including students, particularly students with disabilities, in assessments and the need for assessing students more effectively has created various test administration conditions. For instance, a study conducted by Coniam (2001) to examine the use of audio or video as a form of listening assessment in the certification of English language teachers showed no significant difference between the scores of audio and video groups, however, the conclusion based on overall reactions to test accommodations was that a listening comprehension test should be administered through an audio mode instead of the video mode. Regardless of this conclusion, a major finding in that study included opposing views expressed by the participants: some of the test takers from the audio group would prefer to take the test via video while some others in the video test group expressed that the visuals distracted them while taking the video test.

*Correspondence to: Tufan Adigüzel, Ph.D of Educational Technology, Bahçeşehir Üniversitesi, Department of Computer Education and Instructional Technologies, Çırağan Caddesi, Osmanpaşa Mektebi Sk., No: 4-6, Beşiktaş, 34538 İstanbul, TURKEY
E-mail: tufan.adiguzel@bahcesehir.edu.tr*

State of the literature

- Due to the importance of understanding the language of the content for acquiring science concepts, vocabulary instruction and assessment of science have always been an essential interest of the researchers and practitioners.
- However, the literature presents few approaches and limited measures to assess vocabulary development in science content-area reading.
- Testing accommodations (accommodated and non-accommodated) for a sample of individual students both with disabilities and without disabilities to investigate whether accommodations create inclusive assessments or increase performance have long been studied in the relevant literature, but there is very little research on the use of read-aloud testing accommodation in science.

Contribution of this paper to the literature

- The current study addresses several of the shortcomings (lack of orally accommodated vocabulary assessment) identified in the previous research and focuses on the effect of read-aloud testing accommodation in science assessment tasks.
- This study also provides support for the use of progress monitoring in the content area of science and highlights the importance of investigating the benefits of vocabulary knowledge.
- In addition, the study demonstrates web-based assessment with an audio component as an example of a curriculum-based progress-monitoring tool for elementary science knowledge.

Consequently, having established the fact that teachers view the application of various test accommodations as beneficial such as audio accommodations (or oral administration), computerized procedures, time extension, calculator use, etc., further empirical research on the application of specific test accommodations is highly needed in order to have a better understanding of their effectiveness in learning and to increase the participation in various types of assessments (Bielinski, Thurlow, Ysseldyke, J. Freidebach, & M. Freidebach, 2001; Thompson, Blount, & Thurlow, 2002). This will assist researchers in developing models that will be very helpful to teachers in designing effective instruction methods and assessing students more effectively.

Of particular interest to this study is the use of audio modification in assessments, especially in relation to the fifth grade vocabulary and science. In the course of

vocabulary acquisition, which is influenced both by increment and heterogeneity, there is a potential need for different approaches that would capitalize on these two characteristics. Given the understanding that context does lend power to teaching, it is therefore logical to explore other media through which this objective could be practically achieved. This study delves into a sensory context to complete this objective, and more specifically the influence of auditory contexts in vocabulary assessments.

Theoretical Background

The relevance of vocabulary in reading—especially in reading research and reading instructions—along with the need to teach vocabulary more aggressively, has been emphasized in research (Pearson, Hiebert, & Kamil, 2007). There is no doubt knowledge of vocabulary and reading comprehension are interwoven. It is therefore logical to posit that through both competencies the foundation for learning a specific subject such as science can be laid. Actually, it has been argued that the reasons behind the fusion between the two competencies are centered on the idea of context (Pearson, Hiebert, & Kamil, 2007). Simply put, comprehension is associated with vocabulary skill because comprehension involves wrapping one's head around the idea that context helps in the development of both competencies. Therefore, the adopted approach in vocabulary assessment can be thought to be as important as any of the ones applied in teaching it.

The rapid advancement of technology, including audio, video, computer manipulatives and so on, has played a very active, if not the most active, role in the today's classroom, especially in teaching and assessing students by changing the test administration conditions. The use and relevance of these technological tools have been studied with regard to various classroom concepts and at various levels (Suh & Moyer, 2007; Uribe-Flórez & Wilkins, 2010). The use of audio modification as a technological accommodation is a critical component in vocabulary assessment, and is the primary focus of this study.

Definition and Historical Consideration

Vocabulary is usually referred to as the formalized modification of a natural language to form indexes and thesauri (Buckland, 1999). According to Pearson, Hiebert, & Kamil (2007) what is generally referred to as vocabulary assessment is actually the student's assessment of meaning of words, a concept as old as reading itself.

The history of vocabulary assessment is as old as the time when early tests for intelligence were crafted by Binet and Thursrone (Pearson, Hiebert, & Kamil, 2007)

which came even before the formal measures of reading comprehension. The earliest methods of assessing vocabulary reading comprised of requesting students to either define or describe some intentionally selected words due to the fact that they would be found in the required curriculum. As the world advanced and the English language gained more worldwide acceptance, there became an increasing need to contextualize vocabulary. Thus, standardization of vocabulary assessment began (Pearson, Hiebert, & Kamil, 2007).

Approaches to Teaching Science Vocabulary

Researchers have echoed the significance of tackling students' persistent performance shortcomings by means of compensatory approaches that encompass technology to provide better and easier access to the curricular contents, lessen or eradicate the consequence of a student's disability on achievement in classroom, and thus enhance student learning (Boyle et al., 2003). Assistive technology, such as audios and visuals, has been embraced as a ground for teaching vocabulary based on the fact that it presents an alternate approach to the conventional textbook modification. The alternate approach is often presented via a comprehensive program, which teaches vocabulary in a way that is more detailed than merely listing words. The whole idea of a comprehensive program is to develop both receptive and expressive vocabulary aptitudes—and to constantly transfer words from the receptive domain to the expressive domain. Additionally, the goal includes providing many opportunities for children to gain a deeper understanding of words, as well as the usage of the words they learn (Blachowicz, Fisher & Watts-Taffe, 2005). The comprehensive program is important to avoid the frustration students face in content area classrooms where reading texts contain many unfamiliar vocabulary words (Braun, 2009). It is a known fact that many subjects, particularly in the field of science, contain vocabulary words unfamiliar to students (Chatham County Schools, 2010).

Research has demonstrated that vocabulary comprehension is important in academic achievement, especially achievement in the sciences (Chatham County Schools, 2010). It seems there is enough of a body of literature, or at least close to sufficient literature, on the connection between vocabulary and academic achievement, but research in the context of scientific vocabulary assessment is far scarcer. The current review is therefore an attempt to address this unexplored area.

Approaches to Assessing Science Vocabulary

Generally, the assessment of vocabulary is often fraught with difficulties (National Reading Panel, 2000)—including differentiating between various

categories of vocabulary, accuracy of vocabulary, limitations in regard to the number of new words a learner can accommodate, and so on. This may be the causal factor for the scarce existence of literature on the approaches to assessing science vocabulary. Conceptually, assessment of science vocabulary can be achieved similarly to the approaches used in normal vocabulary assessment.

Research has suggested some science vocabulary alternative assessment approaches. For instance, Dougherty Stahl and Bravo (2010) suggested the use of the Vocabulary Assessment Magazine (VAM) which involves the use of activities that measure students' frequency of science-word use, and was empirically studied by Bravo, Cervetti, Hiebert and Pearson (2008). The VAM has two main parts: (1) the open-ended literacy questions for measuring students' comprehension and prediction abilities, and (2) literacy practices, such as drawing and labeling, for measuring students' science knowledge. Bravo et al. (2008) administered the VAM to 703 second and third grade students to address their vocabulary use and found statistically significant results for English-only students and English-language learners in their sample. However, this may need further investigation because it has only been applied at second- and third-grades. Another example is what Wetzel (2009) calls "Alternative Assessment," which involves the use of a simple rubric to assess a student's mastery level of vocabulary from each science subject area, exposing them to activities—drawing pictures, writing descriptions, or explaining definitions—designed to show their capabilities with scientific language.

The "keyword" approach is one of the most famous and widely researched vocabulary teaching methods (Rodriguez & Sadoski, 2000; Blachowicz, Fisher & Watts-Taffe, 2005). This has formed a theoretical framework for utilizing the keyword method in vocabulary learning and vocabulary assessment, which lies primarily in its ability to encourage verbal connection and visual imagery in the memory process (Blachowicz, Fisher & Watts-Taffe, 2005). The keyword approach utilizes imagery to connect words with their definitions with the aid of audio and visual cues.

With the increased use of approaches used for assessments along with the importance of knowing and boosting the academic progress of students, testing accommodations have long been recognized by educators and policy makers, but have become increasingly important. "Testing accommodations are changes in the way a test is administered or responded to by a student. Testing accommodations are intended to offset distortions in test scores caused by a disability without invalidating or changing what the test measures" (Elliott, Kratochwill, & Schulte, 1999, p. 2).

The most common accommodations found in previous syntheses of research were grouped into the categories of presentation, response, setting, equipment, and timing/scheduling. Utilizing new technological innovations has provided numerous presentation formats to be used. A few of the new presentation formats include braille editions, use of magnifying equipment, large-print editions, oral (read-aloud) administration, signing of directions and interpretation of directions.

The Read Aloud Assessment Accommodation

The “read aloud mode” is a commonly utilized accommodation during which the relevant portions of the test are read aloud to the student via an instructor, computer, video, or audiotape to facilitate the test. Oral presentations typically are not allowed on reading tests or other tests assessing the construct of interest, typically reading comprehension. In contrast, reading the directions and word problems on a math or science test to students who have poor reading skills does not invalidate the test.

Vocabulary fills an important gap in learning, especially learning how to read. As a child begins the process of learning to read, reading vocabulary the child comes across in texts is basically plotted onto the oral vocabulary the child brings to the task. This means, the child is shown how to translate the relatively unacquainted words in print into speech, primarily with the anticipation that the speech arrangements will be easier to understand. An advantage in appreciating text by relating letter-sound information to printed detail only happens if the ensuing oral depiction, or rather representation, is a known, or at least familiar, word in the child’s oral vocabulary. If the subsequent oral vocabulary element is absent from the child’s vocabulary, it will not be better comprehended than it was in print. As a consequence, vocabulary gives such an impression that it occupies very vital middle ground in learning how to read. Oral vocabulary remains a fundamental factor to learning to make the conversion from oral to written representations, whilst reading vocabulary is critical to the comprehensive and/or complete understanding processes of a skilled child (Blachowicz, Fisher & Watts-Taffe, 2005).

Research has confirmed students’ ability to acquire vocabulary while listening to nonfiction read aloud samples that are related to the physical sciences curricular contents (Braun, 2009). Apart from helping students in vocabulary acquisition, the sentiment has been argued by Krashen (2003) that audio mode assists students in getting other linguistic profits from listening and reading, such as better development of grammar and attainment of “planned discourse,” not counting knowledge and pleasure. Research on hypermedia

authoring, which includes audio and video, revealed that hypermedia authoring also had a positive effect on students’ cognition of science concepts, grade level, and on student engagement in, and attitudes of, the direction of vocabulary development activities (Pritchard & O’Hara, 2009).

The past research also includes a variety of experimental studies focusing on K–12 students and presenting test material orally to students in the subject areas of math and science. Of the six studies that focused on oral accommodations on math tests, five (Calhoun, Fuchs, & Hamlett, 2000; Huynh, Meyer, & Gallant, 2004; Johnson, 2000; Tindal, Heath, Hollenbeck, Almond, & Hamiss, 1998; Weston, 2002) found positive effects of oral administration for all students, but one (Helwig & Tindal, 2003) found no gains as a result of read-aloud aid for students either with disabilities or without disabilities. However, studies (Brown & Augustine, 2001; Meloy, Deville, & Frisbie, 2002) focusing on science tests found either no gains for students with disabilities or similar gains for students both with and without disabilities. In summary, the relevant research examines testing accommodations (accommodated and non-accommodated) for a representative sample of individual students with disabilities and without disabilities to investigate whether accommodations remove disability barriers or artificially raise performance. It seems clear oral administration had a positive effect on scores for students with disabilities, but this is not certain for all content areas, including science.

Given the diverse findings of previous research, there is a need for empirical research in science related to effects of oral administration/read-aloud testing accommodation on test scores. The current study addresses several of the shortcomings identified in the previous research and focuses on the development, use, and effect of read-aloud testing accommodation on the scores of students on science assessment tasks. The major objectives of the study were to (a) demonstrate the read-aloud testing accommodations used when assessing students, (b) examine the effect these accommodations have on test results of individual students who did not receive the accommodations and those of the same students who did receive the accommodations in subsequent test items, (c) investigate whether read-aloud accommodations appears to be more beneficial for some at-risk categories, and (d) check the pattern between the frequency of audio use and test scores.

METHOD

Participants

The participants in this study were drawn from a larger sample of fifth grade students who were involved in a funded research project. There were 162 students in the study, who came from 14 schools in 10 different school districts in a south-central state in the United States. A progress-monitoring tool, the Science Key Vocabulary Assessment (SKeVA) (Vannest, Adiguzel, & Parker, 2007) was used over a period of 10 weeks during spring 2009. The SKeVA system was designed especially for students struggling in science by providing teachers for monitoring their students' progress with weekly objective science achievement scores. Therefore, students at risk for failure, as identified by their reading comprehension status, were sought to participate in this study. The several categories of students who were considered at-risk for failure by their teachers compose eighteen percent of the sample, and all students on IEPs (Individualized Education Program) were partially or fully included in the general education classroom (see Table 1 for demographic information about the participants).

Measure

The SKeVA System is a web-based application that has been designed to monitor progress of all elementary students in Science, with the ultimate goal of assisting educators in the preparation of increasing student performance on high stakes assessments and monitoring their students' progress more easily. SKeVA assesses students' knowledge on vocabulary in a fill-in-the blank

Table 1. Student Demographics

Demographic	N
<i>Gender</i>	
Male	68
Female	63
No Data	31
<i>Ethnicity</i>	
White	6
Hispanic	23
African American	4
No Data	129
<i>Categories of At-Risk of School Failure</i>	
Title VI	2
Specific Learning Disability (LD)	10
Resource Room	5
Dyslexia	8
Retained	10
Language (English Language Learner - ELL)	2
Failed Science	7
Other at Risk	108
None	

format tailored to the educational needs of each individual classroom.

Teachers choose the specific vocabulary to be assessed based on specific units of instruction by using the SKeVA system. The content of SKeVA of fifth grade science is broken down into four main content areas: (a) the nature of science, (b) life science, (c) physical science, and (d) earth science. These four areas are represented by a combination of 29 major and 12 subordinate student outcomes, totaling 41 outcomes.

There are no required science texts by the state in which this study took place for fifth grade science. Therefore, there is variation in the science content materials used from school to school. However, schools use science materials which focus on the key student outcomes targeted by the publishers. No test items of SKeVA belong to published materials; all are original creations. Items depend on identifying "key vocabulary" (KV) standing for concepts, principles, and processes which are prominent or central to student learner outcomes. All KV had science-specific definitions. SKeVA has six hundred and seventy KV—about 16 KV for each student outcome in fifth grade. Example KV for the topic, "Inherited and Learned Traits" are: behavior, instinct, adaptation, learned behavior, inherited trait, camouflage, etc. Each item is a single sentence with one blank for an omitted word: the key vocabulary term. The average item sentence is 16 words in length (IQR=12 to 20). For each KV ($N = 670$), there is an average of 3 items (range 1 to 8), to total 2214 items, or nearly 54 items per fifth grade objective. The multiple items per KV include a definition and its important uses. Examples of KV probe items for the KV listed above are:

- (1) *The way an organism acts that is either learned or inherited is called a(n) _____. (Behavior)*
- (2) *A(n) _____ is an action or behavior an animal is born knowing how to do. (Instinct)*
- (3) *The purpose of a(n) _____ is to help an organism survive and reproduce in its environment. (Adaptation)*
- (4) *A bird able to build a nest without learning is an example of a(n) _____. (Instinct)*
- (5) *_____ are actions an organism learns from its parents while living in their environment. (Learned behavior)*
- (6) *_____ are features that an animal is born with, such as hair and eye color. (Inherited traits)*
- (7) *Some animals have _____, or patterns of body color, that help them blend into their surroundings. (Camouflage)*

Efficiency in SKeVA is obtained through computerized test production, administration, scoring, and output. The computer produces equivalent probes based on a repeatable algorithm instantly available for online administration. Probes are self-administered on computer, immediately objectively scored, and students receive immediate textual, numeric and graphic feedback. Textual feedback consists of an item report

QUESTIONS	KEYWORDS
1. The hottest parts of Earth are found around the line called the <input type="text" value="BLACK HOLE"/> .	A. BLACK HOLE
2. When a massive star collapses, it can become a(n) <input type="text"/> .	B. NEPTUNE
3. Similar to the planet Uranus, with a blue and white atmosphere, is the planet <input type="text"/> .	C. EQUATOR
4. Dark spots, or <input type="text" value="SUNSPOTS"/> , are not as hot as the rest of the Sun.	D. ASTEROIDS
5. Scientists sometimes use a meter stick or ruler to <input type="text"/> objects.	E. SUNSPOTS
	F. MEASURE
	SAVE and CONTINUE

Figure 1. Sample student screen on SKeVA

of each test item, the correct answer, and the student's response. The design and format of KV probes considers the following guidelines: (1) to work well on a web-page, (2) to be efficient—to be completed in 10 to 20 minutes, (3) to be followed by low-performing students, and (4) to reduce the amount of skewed data as a result of guessing.

Each probe has 20 items, which are presented in four groupings (separate web-pages) of five items each (see Figure 1). The five items are displayed on the left of the page, and the five related KV response choices together with one distracter are displayed on the right. This provides a 1:1 correspondence between items and KV choices (with the exception of the distracter), so each KV is used only once in each grouping. Selecting from six KV choices yields a low guessing floor (.167) and presenting only 5 items at a time on each page reduces cognitive overload for lower achieving students. Students can respond in one of two ways to each item: drag-and-drop the correct KV or type in a code letter for the correct KV. They also have the opportunity to change their response. As students respond to each item, the answer choice is crossed off in the answer key, which allows students to keep track of their responses. If selected by their teachers, students can also use audio icon next to the items and keywords to listen the oral presentations of such texts professionally narrated by science teachers involved in the project.

On the SKeVA home page, students select the appropriate test to take. To aid in accurate test selection, teachers control which tests are accessible to students. Tests are categorized in three sections on the student home page: upcoming tests, tests in progress, and past tests. While completing tests, students see a timer (if selected by their teachers) indicating the amount of time they spend on the test. If students are not able to complete a test in one sitting, they are able to save their answers after completing one test section and complete the remaining test items at a later time. Once the test is complete, students receive immediate

feedback on their test performance on the item report page. All items on the students' test, the correct answers, and the students' answers are listed, as well as numeric (percent items correct and percent items incorrect) and graphic feedback (a bar graph of the percentage of correct and incorrect answers on the current test and previous tests students take within that testing cycle i.e., on the post-test students can see graphs of pre- and post-test results). Students review the item reports and use the graphs to monitor their progress at any time following test completion. As students complete test probes, student performance is automatically recorded and entered into the report page for teachers and administrators to monitor student performance. Teachers access student performance through individual student and class reports, both of which are accompanied by graphs.

Administration

Students individually visited a classroom or a lab computer. They logged into the SKeVA website using a username only. Passwords were not used to avoid confusion for children and to provide ease of use for teachers. Students completed the assigned SKeVA probe without questions or guidance from their teachers. Students took approximately 5 to 20 minutes to complete the SKeVA probes. Some students in each class had questions regarding which probe they should take on the SKeVA. Some students were confused and had difficulty in finding the correct probe because there was a long list of probes to choose from. Other students had difficulty logging into SKeVA, and could not remember their username because many of the students did not have usernames in the same format. The students who had audio access available used the audio component (with headphones), and some students needed to be reminded by the teacher to use the headphones. The students who had audio access available make the sample of this study. However, there

were some students in the class that had the audio feature available to them on the probes, but they did not use headphones or the audio component. When they had finished, students raised their hands, and the teachers recorded the students' scores to be used for their own reports.

Data Collection

To address the objectives of this study, qualitative and quantitative data were gathered from the participants' test scores, external observer records, and focus group interviews with the participating students. Data collection from the participants' occurred over a 10-week period in the spring 2009. Test scores were obtained from 162 participants who were assigned the audio feature on the SKeVA by their teachers and answered a total of 24,230 items in 1,236 separate probes. Participants used the audio feature for 8,109 of the items, but the same did not use the feature for 16,121 items, even when it was available. Two percentage scores, audio vs. non-audio, were calculated for each participant by dividing a total number of correct items answered with audio or without audio into to the total number of items respectively.

A data collector from the research project observed three classes, approximately 60 students in total, at two different times in order to see how efficient the SKeVA system used the audio support. The same observer did three focus group interviews with a total of seven students. These observation records and interviews were considered supporting data to provide more information about participants' use of the SKeVA system with the audio feature.

Data Analysis

Both qualitative and quantitative data analysis techniques were used to test the effect of read-aloud accommodation and subsequent audio-related hypotheses (Johnson & Onwuegbuzie, 2004). To address the second objective of the study, a paired samples t-test was run to compare the participants' read-aloud accommodated and non-accommodated test scores. Regarding the third objective, the ANOVA was conducted to compare the impact of read-aloud accommodation on scores of at-risk students. Lastly, K-Means analysis was conducted to check the pattern between the amount of audio use and test scores by examining proportions of More Frequent, Intermediate and Less Frequent audio user profiles. K-Means is an unbiased iterative partitioning or clustering method which reduces individual data to two or more profiles based on centroid distances (Blashfield & Aldenderfer, 1988). K-Means analysis iteratively identifies linear or non-linear improvement patterns, and identifies the best

sub-group solution from ANOVA analyses. K-Means was shown to produce sub-group profiles nearly identical to those from the more established, latent-factor technique, Tuckerized Growth Curves (Brossart, Parker, & Willson, 1998). Qualitative analysis was used for non-quantifiable data—interviews and observations—to validate the results from the first three objectives. The analysis techniques were performed separately, but the results were combined at the interpretive stage.

RESULTS

Data from the SKeVA were analyzed to test the differences of two accommodation groups on fifth grade science topics. The descriptive statistics of the participants' scores indicated the participants' level was generally above average. Of mean scores of the 162 participants who used audio presentation for the item contents, 116 were greater than 50 ($M = 81.27$). Similarly, of the mean scores of the same 162 participants who did not use audio presentation for the item contents, 107 were greater than 50 ($M = 75.75$). A paired-samples t-test was conducted to examine students' score differences of the items answered with and without audio support. The scores of students who received read-aloud accommodation appeared significantly higher ($M = 68.47$, $SD = 24.905$) than those of the same students who did not receive audio accommodation ($M = 61.40$, $SD = 23.935$) well beyond $p = .0001$ — $t(161) = 4.878$. These results suggest that special accommodations, specifically read-aloud accommodations, did appear to provide an opportunity for students to demonstrate competency in science without the interference of their disability.

Separate paired-samples t-tests were conducted to examine whether there were any accommodation differences in students' test performances due to at-risk conditions—ELL ($N = 10$), Dyslexia ($N = 8$), LD ($N = 10$) and Retained ($N = 10$). The mean scores for the students with these conditions who used oral administration for item contents ($M_{ELL} = 49.1$, $SD_{ELL} = 32.05$; $M_{DYSLEXIA} = 54.62$, $SD_{DYSLEXIA} = 24.79$; $M_{LD} = 42.71$, $SD_{LD} = 17.6$; $M_{RETAINED} = 44.35$, $SD_{RETAINED} = 25.83$) were higher than their mean scores of non-accommodated items ($M_{ELL} = 46.87$, $SD_{ELL} = 22.99$; $M_{DYSLEXIA} = 40.48$, $SD_{DYSLEXIA} = 19.53$; $M_{LD} = 35.92$, $SD_{LD} = 15.13$; $M_{RETAINED} = 36.53$, $SD_{RETAINED} = 17.6$). However, there were no statistically significant differences in such students' performances due to oral administration: for ELL students, $t(9) = .367$, $p = .722$; for students with dyslexia, $t(7) = 1.682$, $p = .136$; for students with specific learning disability, $t(9) = 1.434$, $p = .185$; and for students who were retained, $t(9) = 1.351$, $p = .21$

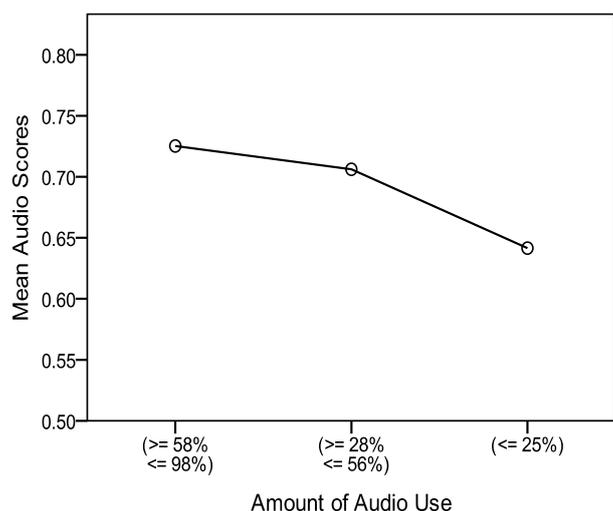


Figure 2. The pattern between amount of use and audio scores

An analysis of 162 students for read-aloud commonalities cannot be easily conducted, so the K-Means analysis clustering technique was used to derive a smaller number of profile types to be able to cluster students by their audio use patterns and then test these patterns with their test scores. A K-Means analysis applied to the 162 students' audio use amounts revealed three profiles, each represented by several students, which were cohesive and quite distinguishable from each other. Also, no students were considered unclassified (i.e. did not fit into any of the clusters). These groups distinguished themselves by the following amounts of audio uses for the item contents: cluster 1 ($\geq 58\%$ and $\leq 98\%$), cluster 2 ($\geq 28\%$ and $\leq 56\%$) and cluster 3 ($\leq 25\%$). Overall, the More Frequent Audio-Use profile (cluster 1) made up $45/162 = 27.78\%$ of the sample, the Intermediate Audio-Use profile (cluster 2) made up $50/162 = 30.86\%$ and the Less Frequent Audio-Use profile (cluster 3) made up $67/162 = 41.36\%$. The most common profile ($N = 67$ students) demonstrated below-average mean scores ($M = 64.15$, $SD = 27.38$). The intermediate-use profile ($N = 50$) included students with higher mean scores ($M = 70.62$, $SD = 25.94$). The less prevalent profile ($n = 45$) was composed of students with the highest mean scores ($M = 72.53$, $SD = 18.6$). These results showed that there is a simple linear trend between the rates of use and the mean scores (see Figure 2). However, the overall profile factor on the participants' scores of items answered within read-aloud accommodation was not statistically significant ($F(2, 159) = 1.81, p > .05$).

The results from focus interviews with seven students and observation records of three classes revealed one overarching theme: it is beneficial to use the audio feature on the test. All participants reported they completed the probes faster and received better scores when using the audio. Most of them indicated that audio accommodation enabled them to have less

questions about the test and to be more independent since they didn't need teacher's assistance or did less teacher's help. One typical comment was also about SKeVA interface saying that it was easy to figure out how to use the audio and find the audio icons on the test.

DISCUSSION

Although sometimes criticized as offering an unfair advantage to students, accommodations can be used to assist students to excel (Lee, Deaktor, Hart, Cuevas, & Enders, 2005). For example, the use of technology in the classroom (Medina-Jerez, Clark, Medina, & Ramirez-Marin, 2007) can assist students to be more academically successful. Utilizing computer technology to assess science vocabulary competency can help to assist students with learning. Technology supported programs can offer multiple forms of representation and communication to all students who are in the process of developing literacy (Shear, 1999). This study demonstrated SKeVA used with an audio component that assessed fifth grade science knowledge, absent of reading ability, and allowed teachers to designate students who might require the reading of items and keywords through the audio system. The application of this component provided information on the effect of an oral testing accommodation on the performance of students.

There were three major findings of this study. The first was students' item scores were significantly higher when the audio mode was used for the items, which is consistent with research completed by Calhoon, Fuchs, and Hamlett (2000); Huynh, Meyer, and Gallant (2004); Johnson (2000); Krashen (2003); Meloy, Deville, and Frisbie (2002); Pritchard and O'Hara (2009); Tindal et al. (1998); Weston (2002). The second major finding was that more frequent use of the read-aloud accommodation led to higher scores on the science test. The third major finding was the read-aloud accommodation on the science test resulted in better performance for an at-risk population of students, but was not statistically significant.

Although there is a significant difference between the scores of the read-aloud accommodated items and non-accommodated items in favor of the former one, non-accommodated mean scores of 55 students were slightly higher than the orally supported ones. One reason for this discrepancy could be that students became more experienced as the test went on and did not need any audio aid. On the other hand, non-accommodated mean scores of more than 65% of the students ($N = 107$) in the study were greater than 50, which might also support the idea that practicing SKeVA can increase students' performance and reduce the frequency of guessing.

This investigation has revealed that progress monitoring in the content area of science, especially for students who initially acquire knowledge at a slower pace, can make reasonable academic gains. Therefore, SKeVA can be used as a progress-monitoring tool to guide instruction and determine when at-risk populations of students are not responding to interventions. This database decision-making process allows for the implementation of targeted interventions to effectively intervene (Klingner & Edwards, 2006). The use of alternate assessments (models, projects, simulations drawings) with various accommodations instead of typical pencil and paper tests allows at-risk populations of students to demonstrate their understanding of concepts in different ways (Medina-Jerez et al., 2007; Olson, Levis, Vann, & Bruna, 2009). The reality is that schools are expected to demonstrate academic mastery through the use of standardized testing. SKeVA, with audio support, can bridge the gap between alternate assessment and standardized testing by providing brief assessments of key vocabulary.

This study also investigated whether read-aloud accommodations appear to be more beneficial for the at-risk population of students. Results revealed that there was a difference between read-aloud accommodated and non-accommodated mean scores of students at-risk of school failure (language, dyslexia, specific learning disability and retained) in favor of the oral group. However, all of the students in this study performed below average in both a read-aloud and non-audio conditions when compared with the other students. This result might prove that when students are provided a read-aloud testing accommodation, their test performance represents a relatively accurate measure of their science skill. In other words, no individual-level measure of reading ability was included in the design. Thus, it was not possible to test whether the variation in accommodation effects was associated with variation in such students' reading skills. Therefore, given the findings of the present study, it would be beneficial to have an emphasis on the appropriate and individualized assignment of accommodations to students (Edgemon, Jablonski, & Lloyd, 2006).

From the perspective of ELL, many ELL may struggle to articulate concepts if they do not have access to English terms (Olson et al., 2009). Furthermore, keywords often provide important labels for the learning of concepts in science. SKeVA, with audio support, can serve to introduce and monitor the learning of specific key vocabulary in science to aid in the development of concept understanding. Moreover, developments to the SKeVA system should include multi-language versions of keywords and items. It has been suggested that providing tests and keywords in multiple languages can benefit ELL. For example,

bilingual students learn new vocabulary more rapidly in their second language when words were first presented in their native language (Pereozzi & Chavez-Shanchez, 1992). This accommodation can help to increase the success of ELL in science education.

A limitation of the present study, similar to that of other studies of oral testing accommodations, is having no measure of reading ability as a factor (reading ability was only decided by teachers of the participants). To measure the usefulness of read-aloud accommodations on a content area test such as science, it might be better to first decide which students have difficulties with comprehension due to their poor decoding and which students have difficulties with comprehension regardless of the representation modes. The distinction can be made by asking students to first complete a reading comprehension test presented both orally and in writing. An additional limitation of the study is the fact that sampling of this study was neither random nor stratified for students. Rather, students who had participated in a larger study on vocabulary assessment in the area of science were solicited. Therefore, there is no certainty for the degree to which the students in this study were representative of other students in their state. The final limitation was that having uncertainty about disability or failure data (the system failure or missing data or meaning that students without disabilities or not at-risks) of 108 students restricted the capacity to run comparative analyses to address the effect of read-aloud accommodations for students without disabilities or not at-risks.

CONCLUSIONS

This study provides support for the use of progress monitoring in the content area of science and highlights the importance of investigating the benefits of oral administration on content area tests. The use of the SKeVA system to monitor the progress of student growth of vocabulary is particularly advantageous due to audio component. This study indicates that read-aloud accommodations have positive effects for certain groups of students. The remaining issues include how to implement different accommodations appropriately, how to identify which accommodations are the best for specific students, and how to develop more flexible tests that would make accommodations unnecessary. Thus, additional research in the area of test accommodations is likely to promote more valid assessment practices in the content areas, specifically in the field of science.

Additionally, recommendations are provided for teachers to maximize effective instruction and assessment for students in inclusive classrooms, and to improve test administration decisions. One method for helping teachers make decisions may be to train them on how to assess performance resulting from substantial

differences when the read-aloud procedure is used. This will allow teachers to assign read-aloud accommodations more strategically for content area tests.

ACKNOWLEDGEMENTS

This research is based in part on the grant project Data to Knowledge, led by Kimberly Vannest and supported by the Texas Education Agency. Any opinions, findings, and recommendations expressed in this publication are those of the author and do not necessarily reflect the views of funding agencies.

REFERENCES

- Bielinski, J., Thurlow, M.L., Ysseldyke, J., Freidebach, J., & Freidebach, M. (2001). *Read-aloud accommodations: Effects on multiple-choice reading and math items* (NCEO Technical Report 31). Minneapolis, MN: University of Minnesota: National Center on Educational Outcomes. Retrieved Jan 4, 2011 from <http://education.umn.edu/nceo/OnlinePubs/Technical31.htm>
- Blachowicz, C.L., Fisher, P.J., & Watts-Taffe, S. (2005). *Integrated vocabulary instruction: Meeting the needs of diverse learners in grades K-5*. Naperville, IL: Learning Point Associates.
- Blashfield, R.K., & Aldenderfer, M.S. (1988). The methods and problems of cluster analysis. In J.R. Nesselrode & R.B. Cattell (Eds.), *International handbook of multivariate experimental psychology* (pp. 447–473). New York: Plenum Press.
- Boyle, E.A., Rosenberg, M.S., Connelly, V.J., Washburn, S.G., Brinckerhoff, L.C., & Banerjee, M. (2003). Effects of audio texts on the acquisition of secondary-level content by students with mild disabilities. *Learn Disability Quarterly*, 26(3), 203-214.
- Braun, P.A. (2009). *The effects of reading nonfiction aloud on the vocabulary acquisition of middle-school students* (Doctoral Dissertation). Retrieved from <http://digitalcommons.nl.edu/diss/9/>
- Bravo, M.A., Cervetti, G.N., Hiebert, E.H., & Pearson, P.D. (2008). From passive to active control of science vocabulary. In *the 56th yearbook of the National Reading Conference* (pp. 122–135). Chicago: National Reading Conference.
- Brossart, D.F., Parker, R., & Willson, V.L. (1998). A comparison of two methods for analyzing longitudinal data: Tuckerized growth curves and an application of K means analysis. *Learning and Individual Differences*, 10, 121-136.
- Brown, P.J., & Augustine, A. (2001, April). *Screen reading software as an assessment accommodation: Implications for instruction and student performance*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Buckland, M. (1999). Vocabulary as a central concept in library and information science. In T. Arpanac et al. (Eds.), *Proceedings of the Third International Conference on Conceptions of Library and Information Science* (pp. 3–12). Zagreb, Croatia: Lokve.
- Calhoun, M.B., Fuchs, L.S., & Hamlett, C.L. (2000, Fall). Effects of computer-based test accommodations on mathematics performance assessments for secondary students with learning disabilities. *Learning Disability Quarterly*, 23(4), 271–282.
- Chatham County Schools. (2010). *Essential science vocabulary kindergarten (2007-2010)*. Savannah, GA: Chatham County.
- Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study. *System* 29, 1-14.
- Dougherty Stahl, K.A., & Bravo, M.A. (2010). Contemporary classroom vocabulary assessments for content areas. *The Reading Teacher*, 63(7), 566-578.
- Edgemon, E., Jablonski, B., & Lloyd, J. (2006). Large-scale assessments: A teacher's guide to making decisions about accommodations. *Teaching Exceptional Children*, 38(3), 6–11.
- Elliott, S.N., Kratochwill, T.R., & Schulte, A.G. (1999). *Assessment accommodations guide*. Monterey, CA: CTB/McGraw Hill.
- Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children*, 69, 211-225.
- Huynh, H., Meyer, J.P., & Gallant, D.J. (2004). Comparability of student performance between regular and oral administrations for a high-stakes mathematics test. *Applied Measurement in Education*, 17, 38–57.
- Johnson, R.B., & Onwuegbuzie, A.J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7) 14-26.
- Johnson, E. (2000). The effects of accommodations on performance assessments. *Remedial and Special Education*, 21, 261–267.
- Klingner, J.K., & Edwards, P. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly*, 41, 108-117.
- Krashen, S.D. (2003). *Explorations in language acquisition and use*. Portsmouth, NH: Heinemann.
- Lee, O., Deaktor, R.A., Hart, J.E., Cuevas, P., & Enders, C. (2005). An instructional interventions' impact on the science and literacy achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching*, 42, 857-887.
- Medina-Jerez, W., Clark, D.B., Medina, A., & Ramirez-Marin, F. (2007). Science for ELLs: Rethinking our approach, practical considerations and support for instruction of English language learners. *The Science Teacher*, 74(3), 52-56.
- 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.
- National Reading Panel. (2000). *Chapter 4, Part I: Vocabulary instruction*. Retrieved Jan 9, 2011, from National Reading Panel: <http://www.nichd.nih.gov/publications/nrp/upload/ch4-i.pdf>
- Olson, J.K., Levis, J.M., Vann, R., & Bruna, K.R. (2009). Enhancing science for ELLs: Science strategies for

- English language learners that benefit all students. *Science and Children*, 46(5), 46-48.
- Pearson, P.D., Hiebert, E.H., & Kamil, M.L. (2007, Apr/May/June). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, 42(2), 282-296.
- Perozzi, J.A., & Chavez-Sanchez, M.L. (1992). The effect of instruction in L1 on receptive acquisition of L2 for bilingual children with language delay. *Language, Speech and Hearing Services in the Schools* 23, 348–352.
- Pritchard, R., & O'Hara, S. (2009). Vocabulary development in the science classroom: Using hypermedia authoring to support english learners. *The Tapestry Journal*, 1(1), 15-29.
- Rodriguez, M. & Sadoski, M. (2000). Effects of rote, context, keyword, and context/keyword methods on retention of vocabulary in EFL classrooms. *Language Learning*, 50(2), 385-412.
- Shear, L. (1999). *Science, language, and deformed frogs: A design framework for linguistically aware science instruction*. Paper presented at the National Association for Research in Science Teaching, Boston, MA.
- Suh, J.M., & Moyer, P.S. (2007). Developing students' representational fluency using virtual and physical algebra balances. *Journal of Computers in Mathematics and Science Teaching*, 26(2), 155-173.
- 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.
- Tindal, G., Heath, B., Hollenbeck, K., Almond, P., & Hamiss, M. (1998). Accommodating students with disabilities on large-scale tests: An experimental study. *Exceptional Children*, 64, 439-450.
- Uribe-Flórez, L.J., & Wilkins, J.L. (2010). Elementary school teachers' manipulative use. *School Sciences & Mathematics*, 110(7), 363-371.
- Vannest, K., Adiguzel, T., & Parker, R.E. (2007). *Science Key Vocabulary Assessment (SKeVA): A web based system for formative assessment and instruction*. (Beta Version) [Web-based application]. College Station, TX: Texas A&M University. Available from http://skeva.tamu.edu/resources/SKeVA_Manual.pdf.
- Weston, T.J. (2002). *The validity of oral accommodation in testing* (NCES 200306). Washington, DC: National Center for Education Statistics.
- Wetzel, D.R. (2009). *Vocabulary building techniques in science: Teaching strategies for making connections with science concepts*. Retrieved Jan 3, 2011, from Suite101.com <http://www.suite101.com/content/vocabulary-building-techniques-in-science-a142993#ixzz1W0idxrvG>

