

Running head: UNIVERSAL DESIGN & ACCOMMODATIONS

UNIVERSAL DESIGN AND ACCOMMODATIONS: ACCESSIBILITY, RELIABILITY,
AND VALIDITY

A DISSERTATION
SUBMITTED TO THE FACULTY
OF
THE GRADUATE SCHOOL OF APPLIED AND PROFESSIONAL PSYCHOLOGY
OF
RUTGERS,
THE STATE UNIVERSITY OF NEW JERSEY
BY
LEAH DEMBITZER
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
OF
DOCTOR OF PSYCHOLOGY

NEW BRUNSWICK, NEW JERSEY

OCTOBER 2016

APPROVED:

Ryan J. Kettler

Linda A. Reddy

Susan G. Forman

DEAN:

Stanley B. Messer

Copyright 2016 by Leah Dembitzer

Abstract

A recent trend in education research and policy has been to employ Universal Design principles for the greatest number of students to equitably and efficiently participate in large-scale assessments. A partnership was made with a computer science team to create a computer-based reading comprehension test with options for audio presentation and extended time testing accommodations. Tests with and without accommodations were given to 131 twelfth grade students in three high schools in New Jersey. Scores on Oral Reading Fluency Curriculum Based Measures (ORF CBMs) determined placement of students in groups of Students Without Functional Impairment in reading fluency (SWOFI, $n = 83$) and Students With Functional Impairment (SWFI, $n = 44$). Accessibility was measured using Test Accessibility and Modification Inventory (TAMI; Beddow, Elliot, & Kettler, 2008) and student evaluations, with TAMI results indicating an adequately accessible test and students rating the overall platform as highly accessible. Reliability change was analyzed using Feldt's test, but yielded mixed results for different forms of the test. The validity results were mixed as well, with data analyses indicating no increase in scores for either group with accommodations, no differential boost, and no relationship between boost and accommodations. SWFIs spent more time and used more accommodations than SWOFIs, indicating that measurement of access skills can predict student need and use of accommodations. The relationship between reading fluency and reading comprehension lessened for SWFIs with accommodations, indicating that the accommodations may have assisted in removing construct-irrelevant skill. The results of this study suggest further inclusion of testing accommodations into general test platforms as an option for all students.

Acknowledgements

Great moments and grand projects are the culmination of many small steps, with constant guidance and assistance from many others required at each juncture. My acknowledgements could have been given at any point during the years of my schooling and will doubtlessly be warranted again many times in the future. The conclusion of my dissertation and doctoral degree provides an opportunity to reflect and recognize all those who have helped me along this journey.

My first thanks go to my advisor and chair, Dr. Ryan Kettler, who creates a supportive and stimulating academic environment full of opportunity for his students. He provided direction and support, answering many lengthy emails with patience and punctiliousness. Aside from outstanding academic knowledge and research skills, I have learned much about writing, advising, and professionalism from him as well.

I would also like to thank my other committee members, Dr. Linda Reddy and Dr. Susan Forman, for their insightful comments and ideas. The Meta-Assessment Team has been a supportive group who have responded to many emails with assistance, advice, samples, and manpower. Thank you for being available and helpful.

Lastly, and most of all, I would like to thank my family; this accomplishment is truly yours. Your strong network of support encircled me through each peak and valley of this journey. All of the favors that you have done for me over these years could fill another dissertation. I can only hope that I can support and help you as we reach many more milestones together.

I wish to acknowledge that much of this research has been funded by the SSSP Dissertation Grant.

TABLE OF CONTENTS

ABSTRACT	II
ACKNOWLEDGEMENTS	III
TABLE OF CONTENTS	IV
LIST OF TABLES	VI
LIST OF FIGURES	VII
INTRODUCTION	1
Testing Accommodations	1
Accessibility	4
Reliability of Scores	6
Validity of Inferences	6
Reading Comprehension and Reading Fluency	10
Reviews of Research on Testing Accommodations	12
Read Aloud Accommodation on a Reading Comprehension Test	14
Extended Time on a High School Reading Test	14
Computer Based Assessment and Universal Design	15
Research Questions and Predictions	19
METHOD	21
Participants	21
Measures	23

Procedures	25
Data Analysis	28
RESULTS	30
Descriptive Data.....	30
Research Question 1 Accessibility.....	34
Research Question 2 Reliability.....	38
Research Question 3 Validity	40
DISCUSSION	42
Accessibility.....	44
Reliability.....	47
Validity	48
Summary of Findings.....	50
Limitations	51
Future Research	51
Conclusions.....	52
REFERENCES	54
APPENDIX A.....	63
APPENDIX B	64
APPENDIX C	65
APPENDIX D.....	84

LIST OF TABLES

Table	Page #
1. Commonly Awarded Testing Accommodations.....	2
2. Demographic Characteristics of Total Sample	22
3. Data Analyses	30
4. Means and Standard Deviations of Scores.....	33
5. TAMI ARM Results	34
6. Coefficient Alpha for Reading Comprehension Test Full Sample	39
7. Coefficient Alpha for Reading Comprehension Test by Group and Form.....	39
8. Mean Comparisons	40
9. Correlations Between Reading Fluency and Reading Comprehension	41
10. Correlations.....	42

LIST OF FIGURES

Figure	Page #
1. Sequential evidence evaluated for universally designed testing accommodations.....	4
2. Sequential evaluation of valid testing accommodations.....	8
3. Constructs of UD&A Study	20
4. Screenshot of final computer program design	26
5. UD&A Study design.....	28
6. Frequency of CBM scores	31
7. Major findings from UD&A qualitative analysis	46

Introduction

The Individuals with Disabilities Education Act of 1997 and subsequent Individuals with Disabilities Education Improvement Act of 2004 require states to provide equitable assessment for students with disabilities (SWDs; IDEA, 1997; IDEIA, 2004). States are tasked with including SWDs in their large scale assessments in order to ensure that these students will be subject to high expectations and educational opportunity (Fuchs, Fuchs, & Capizzi, 2005; Kettler & Elliott, 2010). Individualized Education Program (IEP) teams provide SWDs with accommodations to allow them to demonstrate their knowledge on the assessments. In accordance with these policies, there has been a steep increase in the volume of testing accommodations offered to and requested by SWDs, as well as in research addressing these trends (Cormier, Altman, Shyyan, & Thurlow, 2010). Another newer trend in assessment has been to design large scale assessments according to the principles of Universal Design (UD; Center for Universal Design, 1997). The purpose of this study, hereafter referred to as the Universal Design and Accommodations (UD&A) Study, is to add to this growing base of knowledge by studying the accessibility, reliability, and validity of a computer based test with universally designed accommodations.

Testing Accommodations

Testing accommodations are changes made to a test's setting, presentation, and/or response modality to allow students to fairly demonstrate their knowledge (Hollenbeck, 2005). According to Fuchs and Fuchs (2001), testing accommodations "level the playing field" for SWDs by minimizing the effects of their disabilities and ensuring that they can demonstrate their competence. Fuchs et al. (2005) have classified testing accommodations by setting, timing, format, and supports. Some examples of testing accommodations given in most states are extra

time, read-aloud/audio presentation, response to scribe, and enlarged test format (Bolt & Thurlow, 2004; Christenson et al., 2008; Fuchs et al., 2005). Table 1 lists common accommodations allowed and awarded in many states.

Table 1

Commonly Awarded Testing Accommodations

Categories:

Setting

- Separate Testing Location
- Small group administration
- Individual administration

Timing

- Extended Time
- Extra Breaks

Test Format

- Braille or large print
- Audio presentation
- Sign directions
- Scribe dictation

Supports

- Communication devices
- Calculator for math tests
- Computer for writing

Note. Compiled from Fuchs et al., 2005 and Christenson et al., 2008.

Two of the most common testing accommodations offered to students with a variety of disabilities are audio presentation and extra time (Cormier et al., 2010). However, most students who need accommodations receive multiple accommodations (Elliott, Kratochwill, & McKevitt, 2001). Some researchers view this as a rationale to study accommodations in packages (Elliot et al., 2001), while others (Fuchs et al., 2005) believe it is more appropriate to first establish the effectiveness of individual accommodations before combining them.

Because testing accommodations are thought to change students' scores, they are often sought by students and can be controversial (Thurlow et al., 2000). CBT/McGraw Hill (2000; 2005) have created three categories of testing accommodations for academic tests with levels of caution needed for interpretation of scores. A "Category 1" accommodation as defined by CBT/McGraw Hill is not expected to influence student performance in a way that would invalidate the standard interpretation of the test scores. For example, a student taking a test alone instead of with a group should not be a concern for test scores. A "Category 2" accommodation may affect student performance and CBT/McGraw Hill recommends considering the accommodation used when interpreting the scores. Extra time has been classified as a Category 2 accommodation. Lastly, a "Category 3" accommodation is thought to change the construct being studied and results from such an accommodation should be interpreted with caution. As an example, allowing use of a calculator for a computation test should be a concern for score interpretation. CTB/McGraw Hill has also considered audio presentation for a reading (decoding) test to be a Category 3 accommodation, and audio presentation of other subject matters to be a Category 2 accommodation (CTB/McGraw Hill, 2000; 2005).

As per the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014), evidence of reliability and validity is necessary for all educational assessments. When

adding testing accommodations, changes are made in the standardization of the test. Therefore, evidence of reliability and validity is needed to support the use of the testing accommodations (Kettler, 2011). Kettler and Feeney-Kettler (2011) proposed a linear framework to evaluate assessments based on the guidelines provided by the *Standards*. They proposed that the accessibility of the measure must be evaluated, then reliability of scores, and finally the validity of the inferences. The UD&A Study will focus on accessibility, reliability, and validity in evaluating universally designed testing accommodations. Figure 1 depicts the sequential steps evaluated in the UD&A Study.



Figure 1. Sequential evidence evaluated for universally designed testing accommodations.

As outlined in Figure 1, if a test is inaccessible, it will not be reliable. Further, acceptable reliability of scores is a prerequisite for investigating the validity of the inferences.

Accessibility

Beddow, Kurz, and Frey (2011) describe test accessibility as the degree to which a test allows examinees to demonstrate their knowledge on a construct. Beddow (2012) further explains that individual test items need to be accessible for an overall score to facilitate valid inferences and decisions. An accessible test item will allow a student entry to the target content being measured without barriers. For example, a multiple choice test item must be visible for its content to be accessible (Beddow, 2012). Accessibility can be depicted as an interaction between the characteristics of a test-taker and features of a test item. When the features of a test item

require certain test-taker characteristics (such as sensory perception or reception) that are not intended to be measured, the test is less accessible for students who have impairments in those areas. This also causes the inferences drawn from the scores to be less valid, as a source for the score variance is not performance on the construct being measured; rather, it is the access skills (Beddow, 2012). Accessibility can also be conceptualized as ensuring that test items are not biased toward any subgroup of students, including students with disabilities (Thompson, Johnstone, and Thurlow, 2002).

Test accessibility can be maximized through test-taking skills training (Kettler, Braden, & Beddow, 2011), through accommodations based on deficits in access skills, and through good item writing and revision (Beddow, 2012). The UD&A Study will focus on using accommodations to improve access, as well as evaluating the accessibility of a computer based testing system. The Test Accessibility and Modification Inventory (TAMI, Beddow, Elliott, & Kettler, 2009) is a tool that can be used to evaluate test items for accessibility. Developed based on principles of Universal Design (UD), test accessibility, Cognitive Load Theory, and fairness, TAMI allows items to be rated for accessibility based on aspects of the item. The stimulus, question, visuals, answer choices, and the layout of items are systematically evaluated to maximize accessibility (Beddow et al., 2009).

Student perception of utility of accommodations can also provide information about response processes, perceived needs, and student judgment. Rogers, Christian, and Thurlow (2012) reviewed testing accommodations literature and found a larger emphasis on studies focusing on student perceptions in 2009-2010 than had been previously found. Student perception has added important aspects to accessibility research. For instance, Higgins, Fedorchak, and Katz (2012) have found that too many tools and supports available during an

assessment can be confusing to students. In a study using think-aloud protocols as students worked through test items and evaluations at the end, Roach, Beddow, Kurz, Kettler, & Elliot (2010) found that students' favorable perceptions of testing modifications did not necessarily match the impact of the modification on students' scores. Student voices on the clarity, fairness, and accessibility of an assessment tool can assist the development of appropriate testing practices (Ketterlin-Geller, 2005; Roach & Beddow, 2011), although positive student responses may indicate psychological comfort rather than score gains (Roach et al., 2010).

Reliability of Scores

Reliability is the prerequisite to investigating validity (Kaplan, 1987). Despite the necessity of measuring reliability as enumerated by the *Standards* (AERA, APA, & NCME, 2014), most research on testing accommodations has not focused on reliability. Kettler et al. (2011) have studied the reliability of modified test items using a meta-analytic approach and adjusting coefficient alphas to compare them. They found significant and relatively minor differences in reliability when modifying items. Since the error in scores resulting from a student's functional impairment is likely systemic error rather than random error, it is possible that reliability would not change substantially. However, evidence of improvement in reliability for SWDs supports the premise that an accommodated assessment decreases error in scores.

Validity of Inferences

According to the *Standards*, an assessment should measure the target construct meant to be tested and eliminate irrelevant factors to the greatest extent possible (AERA, APA, & NCME, 2014). Tests have a target construct (e.g. reading skills, math computational skills) and "access skills" may be needed for a student to demonstrate his/her knowledge on the target construct of the test. The test is not meant to measure access skills, yet the access skills are often the means

through which students can demonstrate competency on a task. For example, vision is an access skill for a test on reading fluency and motor control is an access skill for an essay task (Beddow, 2012; Beddow et al., 2011; Kettler, 2012). When a student lacks one or more access skills needed to demonstrate knowledge of a test's construct, this can be conceptualized as a "functional impairment" in the access skill. Although a test could be designed for maximal accessibility, individual student characteristics and functional impairments create a barrier to the target skills meant to be assessed. Understanding a student's functional impairment can assist in choosing effective testing accommodations, more so than knowing a student's disability category (e.g. Learning Disability, Other Health Impaired). Targeting a specific functional impairment allows a student to overcome any access barriers and demonstrate his/her knowledge on the task (Kettler, 2012).

The goal of academic assessment is to obtain a valid measure of a student's functioning in an achievement domain. If a functional impairment in an access skill is causing a student to do poorly on a test, the validity of the inferences drawn from the test is compromised. Testing accommodations selected to overcome a functional impairment can increase the validity of the inferences drawn from the test (Kettler, 2012). However, if an accommodation will impact the construct measured, it can negatively affect the validity of the inferences drawn from the test scores (CTB/McGraw Hill 2000; 2005). To continue the earlier example, the eyeglasses accommodating a visual functional impairment would be wholly inappropriate on a test measuring visual acuity.

Hollenbeck (2002) delineated key components for valid testing accommodations. He stated that testing accommodations must have the attribute of unchanged constructs, in that they assist with the access skills and do not assist with the construct under measurement. In addition,

testing accommodations should be based on individual needs of students, should have differential effects for students who need the accommodations versus students who do not, and should lead to scores that can be used to make valid inferences and decisions. Kettler (2012) organized these attributes of testing accommodations linearly. Considering individual needs will lead to unchanged constructs on the test. Whether constructs are unchanged can be evidenced by differential effects, and if established this can lead to sameness of inferences that can be drawn from the test (Kettler, 2012). According to Hollenbeck (2005), IEP teams should first identify the purpose of the test and the consequence of the scores. They can then decide whether the proposed accommodations change the construct, are based on need, produce a differential outcome, and lead to appropriate inferences (Hollenbeck, 2002, 2005). Figure 2 depicts the attributes of testing accommodations used in the UD&A Study.

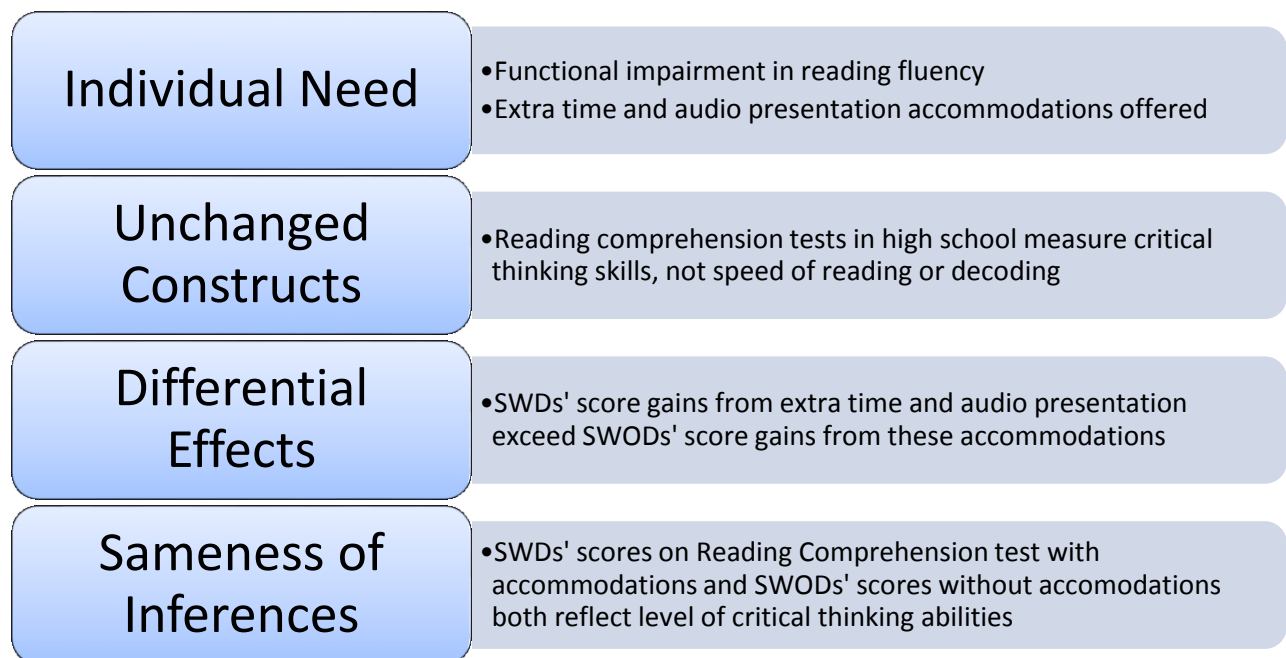


Figure 2. Sequential evaluation of valid testing accommodations. The left column contains the attributes needed for valid testing accommodations (Kettler, 2012), and the right column describes the application of these concepts in the UD&A Study.

As developed in Figure 2, the UD&A Study focuses on the individual needs of a student who has a functional impairment in reading fluency. In considering the students' needs on a test of reading comprehension, it is probable that students with a functional impairment will need extra time and/or audio presentation accommodations to demonstrate the critical thinking skills measured on these high school tests. Test scores as evidence of differential effects were gathered in this study, analyzing whether the construct was unchanged by allowing for these accommodations. The purpose of these accommodations is to allow for inferences about critical thinking to be drawn from a reading comprehension test for both students with and without a functional impairment.

Fuchs and Fuchs (2001) developed *The Dynamic Assessment of Test Accommodations* (DATA) as a tool for teachers and IEP teams to use when choosing accommodations for students (Fuchs et al., 2000, Fuchs & Fuchs 2001). Using DATA, teachers administer brief tests with and without accommodations and compare the results to normative data. The scores are used to determine whether the accommodation meets the criteria of differential effects (Hollenbeck, 2002), also referred to as differential boost (Fuchs & Fuchs, 2001).

A differential boost should be observed when an accommodation is appropriate, with SWDs benefitting more from the accommodation than students without disabilities (SWODs) do (Fuchs & Fuchs, 2001). Although students without functional impairments may gain some benefit from testing accommodations as well, Fuchs and Fuchs (2001) considered it fair to allow SWDs to use these accommodations if a differential boost is found. Sireci et al. (2005) originally suggested a stricter interaction hypothesis of testing accommodations. According to the interaction hypothesis, valid testing accommodations should improve scores for SWDs and not at all for SWODs. In their review of testing accommodation research, they revised their definition

of the interaction hypothesis to make it more similar to the differential boost criteria, due to the lack of support found for their original level of differential effects. Sireci et al. (2005) explained that the interaction hypothesis may not have been found in research due to possible problems with the fairness of assessments toward SWODs.

Reading Comprehension and Reading Fluency

Reading comprehension has been defined as locating, recalling, integrating, interpreting, critiquing, and evaluating read material (NAEP, 2013; Pearson & Hamm, 2005). Reading comprehension is a skill necessary for academic success in many areas (Pearson & Hamm, 2005). Since reading comprehension involves higher level thinking such as synthesizing information and thinking critically, a lack of reading fluency could be an access barrier toward demonstrating these higher level thinking skills (Fuchs et al., 2001; Rasinski et al., 2005). Rasinski et al. (2005) defined reading fluency as decoding automatically and effortlessly. If a student has a functional impairment in reading fluency, it is difficult to demonstrate his/her full abilities in the critical thinking aspect of reading comprehension. For example, a student with dyslexia may have mastered the higher level thinking required to compare and contrast poems. However, due to an impairment in reading fluency, the validity of inferences regarding his/her capabilities in this area will be compromised.

Poor reading comprehension is often associated with poor fluency (Pinnell et al., 1995; Shinn et al., 1992). There are those who argue that reading fluency is a component of skilled reading, and that the reading comprehension construct should reflect this point (Sarroub & Pearson, 1998). For this definition of reading comprehension, a testing accommodation to minimize the effects of a functional impairment in reading fluency would be inappropriate, as reading fluency is conceptualized as part of the target skill.

This study will instead define the construct of reading comprehension in twelfth grade as a critical thinking test. The rationale for this follows a research base on reading fluency progression throughout the school years, a theoretical framework about the purpose of assessment, as well as the stated objectives of the reading comprehension questions used from the NAEP data base. Fuchs et al. (1993) cautioned that although oral reading fluency may be a strong indicator of reading competence in lower grades, more research is needed to understand the construct of reading competence in high school. They found that high school reading development is focused on literary analysis of narratives, as well as processing different kinds of texts (Fuchs et al., 1993; Fuchs et al., 2001). It is apparent that the construct of reading comprehension in high school is not adequately defined, and it is unclear what role actual reading of text plays in the constructs targeted in high school. In their review of reading comprehension tests, Keenan et al. (2008) found that reading comprehension tests differ with regard to their stated constructs, with some more heavily impacted by decoding skill and others by comprehension. In the upper high school grades, the tests reviewed seem to differ less, indicating a lesser impact of decoding skills on the construct (Keenan et al., 2008), and more of a focus on critical thinking. The researchers have said that a clarification of exactly what is meant to be measured on reading comprehension tests can only benefit students, teachers, and researchers.

From a theoretical perspective, assessments should follow instruction and inform future instructional decisions. Some reading comprehension tests measure reading fluency, a skill not taught in high school. Further, it is difficult to see how the results of such an assessment could be useful for instructional planning. Low scores on a reading comprehension test with these mixed constructs may indicate either poor fluency or poor grasp of critical thinking skills, sowing

confusion for instructional planning. It is more beneficial to study each piece alone to guide teacher interventions (Rasinski et al., 2005). If the goal of assessment is to identify students with poor reading fluency for instructional or classification purposes, a more efficient way to do this is to employ specific reading fluency tests, such as three minute long Curriculum Based Measures (CBMs).

The questions for the reading comprehension/critical thinking assessment were taken from the National Association for Educational Progress (NAEP). In their definition, NAEP states that “reading is a dynamic cognitive process that allows students to understand written text, develop and interpret meaning, and use meaning as appropriate to the type of text, purpose and situation,” (NAEP, 2013), indicating a mix between reading skills and comprehension skills. However, NAEP defines cognitive targets for the assessment: locate and recall, integrate and interpret, critique and evaluate, and a vocabulary component. It appears that none of these constructs are related to reading as a fluent process, and all are solely focused on the critical thinking of comprehension. Therefore, this study conceptualizes a twelfth grade reading comprehension assessment as having a target construct of critical thinking and synthesis, with reading fluency an access skill needed to display knowledge on the construct. This narrow definition of the construct assists in the generalizability of the findings for other content areas of study (e.g. science, history) that require critical thinking, but are not meant to measure reading abilities.

Reviews of Research on Testing Accommodations

Fuchs et al. (2005) performed a review of the literature and found that extra time increased scores for students with and without disabilities. They concluded that the accommodation of extra time does not have a differential boost. Fuchs et al. stated that students

with trouble reading may differentially benefit from extra time on a math tests, since this will allow the students to demonstrate their knowledge. They found mixed results for read-aloud accommodations.

Sireci, Scarpati and Li (2005) reviewed 59 studies on testing accommodations. They found that the most studied categories were oral presentation and extended time, and often these two accommodations were given together. They noted that the high school grades were not frequently studied. Most studies were done using math tests and a few were done on reading tests. In reviewing studies of extended time on reading tests, Sireci et al. (2005) found very mixed results. In some studies evidence of an interaction hypothesis/differential boost was found, but in some studies SWODs also received a substantial boost from the extra time. Sireci et al. hypothesized that when a reading test is “speeded,” that is, speed of reading and processing is necessary to complete the task, no differential boost is found (Sireci et al., 2005). Sireci et al. did not find any evidence for the interaction hypothesis for oral presentation on reading tests.

Cormier et al. (2010) reviewed research on the effects of testing accommodations based on 40 studies done in 2007-2008. They found mixed results for audio presentation, mixed results for extended time, and mixed results for aggregated accommodations. They also did not find many studies focusing on the high school population.

Rogers, Christian and Thurlow (2012) reviewed the research done on testing accommodations based on studies done in 2009-2010 and again found mixed results. While some studies showed a benefit for SWDs using extra time, some showed no benefit. In addition, some studies showed differential effects for audio presentation accommodations on a reading test, while others showed similar benefit for all students.

Read Aloud Accommodation on a Reading Comprehension Test

Meloy, Deville, and Frisbie (2002) studied the effects of a read aloud accommodation on tests of science, math, and reading comprehension. Their sample consisted of middle school students with learning disabilities in reading and students without disabilities. Results found no differential boost in any of the three subjects; students with and without learning disabilities both benefitted. Meloy et al. hypothesized that a read aloud accommodation would invalidate a reading comprehension test because SWDs and SWODs will benefit equally from it, but that a differential boost would be found for science and math. Although they did not find a differential boost for this accommodation on the reading comprehension test, no differential boost was found in the other subjects either. Meloy et al. indicated that a lack of student input in pacing and usage of the read aloud may have led to these results, as students were forced to slow their pace and were directed to focus. The UD&A Study will attempt to answer this question carefully by eradicating the confound of teacher direction and allowing students a choice of when to have the test read aloud.

Extended Time on a High School Reading Test

Lewandowski, Lovett, and Rogers (2008) tested the effect of extended time on students classified with LD in the area of reading. SWDs and SWODs were tested using the reading comprehension subtest of the Nelson Denny Reading Test (NDRT; Brown, Fishco, & Hanna, 1993) under time constraint. The researchers shortened the standard time allotment for the reading comprehension test to control for ceiling effects and all students were tested using the shortened standard time, switching to colored pens at the end to continue into extended time. Students were also administered a reading fluency subtest from the Woodcock Johnson Tests of Achievement. Items correct, items attempted, and percent correct of the reading comprehension

test were recorded for both conditions and analyzed. The scores of the LD group were significantly lower on the reading fluency subtest. A differential boost of results was not found, as SWODs benefitted more from the accommodation than SWDs. However, the SWDs were found to be able to attempt as many items in extended time as SWODs were able to attempt in the standard time; the researcher hypothesized that this increased the access of the test for SWDs. Lewandowski et al. hypothesized that controlling for ceiling effects by making the tests more speeded may have prevented a differential boost from being found. They concluded by suggesting that tests should minimize the timing aspect when speed is not the construct studied. This way, all students will be able to access the test items. The UD&A Study was not speeded although a time limit was provided, and the extended time accommodation will be studied for its effect in allowing SWDs to access as much test content as SWODs, and thereby display their true knowledge. Lewandowski et al. (2008) suggest using computer based assessment and universally designed tests to achieve this goal.

Computer Based Assessment and Universal Design

The current study will add to this important and developing body of knowledge by applying principles of UD to assessment to create a more efficient and equitable way of testing SWDs. UD was conceived in architecture to design inexpensive and aesthetically pleasing physical accommodations for people with disabilities (Center for Universal Design, 1997; Story, Mueller, & Mace, 1998). For example, one UD principle is Equitable Use, as defined by providing same means of use for all users, avoiding stigma and segregation, and offering an appealing design with privacy included. The classic example of UD is the sloping curb at the end of a sidewalk, a necessary accommodation for wheelchair users made universally available.

These principles extend to many other fields, including educational assessment. When applying the principles of UD to assessment (UDA), the goal is to provide an equitable and efficient way for the greatest number of students to participate in the assessment (Thompson, Johnstone, & Thurlow, 2002). Thompson et al. define UDA as an inclusive assessment system that is “designed and developed from the beginning to allow participation of the widest possible range of students, in a way that results in valid inferences about performance for all students who participate in the assessment” (Thompson et al., pp. 3). They identified seven elements of UDA based on the general principles of UD. For example, Thompson et al.’s identified element of “inclusive assessment population” reflects the UD principle of Equitable Use (2002). Table 2 depicts key elements in UDA.

According to Thompson et al. (2002), UDA includes the following elements: inclusive assessment population, clearly defined constructs, accessible items, amenable to accommodations, simple and clear instructions, maximum readability, and maximum legibility. Computer-based tests offer an opportunity to develop universally designed assessments because accommodations can be incorporated into the general test design (Dolan et al., 2005; Ketterlin-Geller, 2005; Russell, Hoffman, & Higgins, 2009).

Table 2

Elements of Universal Design in Assessment

Element	Description
Inclusive Assessment Population	Assessments should be designed for students with differing cognitive abilities, cultures, and linguistic backgrounds. Different formats should be used and allowed to accommodate different learning styles.
Precisely Defined Constructs	The construct intended to be measured by the test should be clearly delineated in order to facilitate appropriate inferences.
Accessible, Non-Biased Items	Items of a test should be non-biased to subgroups of students, and should be field tested with different groups to ensure that they are universally accessible.
Amenable to Accommodations	Tests should be able to be accommodated to address the needs of specific students.
Simple, Clear, and Intuitive Instructions and Procedures	Instructions must be easy for students to comprehend, regardless of their level of experience and knowledge.
Maximum Readability and Comprehensibility	Questions should be worded to allow students to demonstrate target knowledge, rather than abilities in comprehending the demand of the question.
Maximum Legibility	All text, tables, figures, and response formats should be as legible as possible, with attention paid to contrasts, type size, and spacing.

Note. Elements and descriptions compiled from Thompson et al. (2002).

Ketterlin-Geller (2005) found that UDA can be conceptualized as a fixed system or an adaptable testing environment. She recommends embedding general supports to all students as well as specific accommodations to maximize performance of students with disabilities. Ketterlin-Geller (2005) described how a UDA can be designed by following a clear procedure. Definition of the construct is the first step, followed by conceptualizing the test construction and choosing the testing platform and format that would fit the purpose of the assessment. Supplemental accommodations are embedded based on specific needs of students, but the researcher did not clarify whether they should be universally accessible, or made available for those who need it. The final step is reviewing test items to ensure that they are free of bias, and receiving feedback from members of the educational community on the ease of use of the testing platform (Ketterlin-Geller, 2005).

In reviewing the literature on testing accommodations, Sireci et al. (2005) recommend that assessments turn to UD to decrease the need for separate testing accommodations. With the new wave of large scale assessments (e.g. Partnership for Assessment of Readiness for College and Careers, PARCC; Smarter Balanced), there has been an increased emphasis on UDA. PARCC and Smarter Balanced are multi-state assessment consortia funded by the federal Race to the Top grant competition (King, 2011). The consortia were tasked with creating assessments that are linked to the Common Core Standards and are credibly tied to college readiness goals. The new assessments became operational in the 2014-2015 school year and the data is pending on their merits and effectiveness. Both PARCC and Smarter Balanced were created according to principles of UD and contain many embedded supports. Both testing systems allow accessibility features and universal tools for all students, specialized additional supports that any student can

request, and a last tier of accommodations for students with IEPs (PARCC, 2013; Smarter Balanced, 2013).

Research Questions and Predictions

The UD&A Study answered three primary research questions:

1. What is the perceived accessibility of a universally designed assessment with choices of accommodations? Accessibility will be evaluated through (a) TAMI (Beddow et al., 2009) ratings, (b) student selection to use accommodations, and (c) qualitative data from student surveys.
2. To what degree will the accommodations of extra time and audio presentation affect the reliability of scores? Reliability will be evaluated by comparing coefficient alphas across groups and conditions.
3. To what degree will the accommodations affect the validity of the inferences that can be drawn from tests scores? Improved validity will be evaluated through (a) differential boost between group and condition, (b) relationship between number of accommodations accessed and boosts in scores, and (c) relationship between reading fluency and reading comprehension for SWODs across conditions.

It was hypothesized that a universally designed assessment with testing accommodations would be highly accessible as measured using TAMI (Beddow et al., 2009). Based on Higgins et al. (2012), it was hypothesized that students with a functional impairment in reading fluency would select and use available testing accommodations, but that students with no functional impairment will not. Reliability was expected to be marginally but significantly improved for SWDs using testing accommodations. Allowing a choice of accommodations was expected to

improve the validity of inferences that could be drawn from test scores. A differential boost was expected to be found. A relationship between accommodations accessed and the boost in scores for SWDs was expected as well. A very large relationship (around .80) was expected between reading fluency and reading comprehension scores for SWDs in the non-accommodated condition, as found by Shinn et al. (1992). The relationship should be lessened when accommodations are offered. Figure 3 depicts the constructs researched in the UD&A Study.

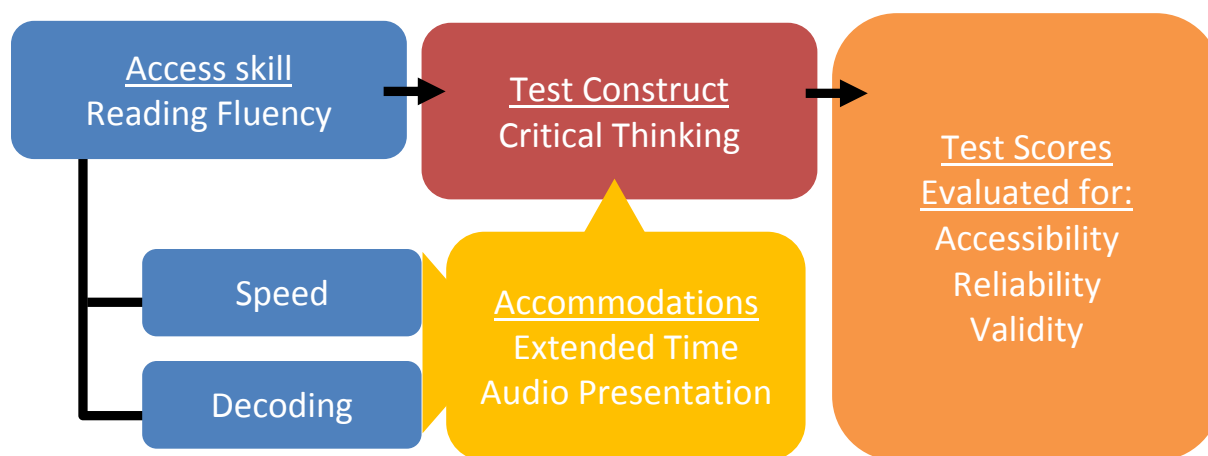


Figure 3. Constructs of UD&A Study.

As portrayed in Figure 3, reading fluency comprises speed and accuracy of decoding and is an access skill needed to display knowledge on a reading comprehension test. Reading comprehension is the academic subject tested. The testing accommodations of extended time and audio presentation are meant to allow students to overcome deficits in the access skill of reading fluency, specifically targeting lagging speed and decoding trouble. The goal of these accommodations is to provide a more accessible, reliable, and valid assessment of reading comprehension, and evidence of these gains is necessary to support the use of testing accommodations in this context.

Method

Participants

The sample included 131 high school students in twelfth grade from three high schools in New Jersey; one was a public school and two were private all-girls schools. The sample was majority female and ethnically diverse. Table 3 depicts the demographic characteristics of the students in the sample.

Table 3

Demographic Characteristics of Total Sample

Characteristic	<i>n</i>	%
Gender		
Female	98	75
Male	25	19
Not Identified	8	6
Ethnicity		
African American	43	33
Asian American/ Pacific Islander	2	2
European American	60	46
Latino/a American	10	8
Native American	0	0
Other/Not Identified	16	12
Classification		
General Education Student	117	89
Special Education Student	5	4
Not Identified	1	1
Reading Fluency		
Functional Impairment	44	34
At-Risk	4	3
No Functional Impairment	83	63
Total	131	100

Note. Functional Impairment was defined as having reading fluency scores one-half standard deviation below the mean, No Functional Impairment was defined by reading fluency scores within measurement error of the mean or higher, At-Risk fell into neither of those groups. All other variables were self-reported.

Measures

The computer based system was evaluated using the Test Accessibility and Modification Inventory (TAMI, Beddow et al., 2009). Participants completed three curriculum based measures (CBMs) for Reading Fluency and two forms of a reading comprehension test. Further, participants completed an evaluation form and a demographic questionnaire.

Test Accessibility and Modification Inventory. The Test Accessibility and Modification Inventory (TAMI; Beddow et al., 2009) was created to evaluate test questions, materials, formats, and stimuli to ensure maximal accessibility. Created from theories of UD, accessibility theory, fairness, and Cognitive Load Theory, the development of TAMI was also aided by practical research on testing accommodations, item writing, and item modification (Elliott et al., 2010). TAMI includes categories of Passage/Stimulus, Item Stem, Visuals, Answer Choices, Page and Layout, and Fairness, yielding ratings for each category as well as a Total Item Accessibility Rating.

Curriculum Based Measures. The Curriculum Based Measures (CBMs) were created by taking three grade level passages from *Bader Reading and Language Inventory, Fifth Edition* (Bader, 2005) and formatting it through Intervention Central's CBM generator to make Oral Reading Fluency (ORF) CBMs. According to Rasinski (2005), reading fluency and performance can be adequately measured and predicted by ORF CBMs. Grade level passages were used,

according to the recommendations of Hosp et al. (2007). The measures were used to classify participants into groups of students with and without a functional impairment in reading fluency.

Reading CBMs provide reliable scores from which valid inferences about reading impairments can be drawn (Hosp et al., 2007). In their review of the literature, Shinn et al. (1992) have found that ORF CBMs are generally reliable in measuring students' fluency. They reviewed numerous studies that have found correlations between ORF CBMs and norm-referenced reading tests ranging from .60 to .90, with most around .80. These large and very large correlations (as classified by Cohen, 1992, and Hopkins, 2001) can be interpreted as ORF CBMs and reading tests representing highly related and possibly overlapping constructs (Kettler et al., 2013), indicating criterion validity of the measure.

Reading Comprehension Test. The Reading Comprehension Test was a computer program in two forms; Form A and Form B. Each form included two passages and 26 multiple choice reading comprehension questions. The passages and questions were taken from the National Assessment of Educational Progress public access item bank. In the NAEP, each student completes two blocks of items, with 25 minutes per block. Each block in the NAEP consists of one passage with multiple choice questions and constructed response questions.

This study only included the multiple choice questions from each passage and the constructed response questions were rewritten as multiple choice questions. Participants completed both forms in two separate timed sessions. During one administration the test included options to have the reading passage and questions read aloud and to allow time and a half.

Demographic Questionnaire. Participants completed a demographic questionnaire on the computer which included information such as gender, age, placement, and ethnic/racial background. The demographic questionnaire is included in Appendix A.

Evaluation Survey. The evaluation survey was used to gain qualitative and quantitative information regarding the subjective experience of the participants in completing the reading comprehension test. Questions addressed organization, clarity, and ease of use of the testing platform and inquired whether the accommodations offered were helpful. This survey was completed on the computer as well. The evaluation questions used are presented in Appendix B.

Procedures

Stage 1: Developing Testing Platform. A partnership was made with a computer scientist to develop a universally designed testing platform. The test was created following Schneiderman's (1997) model for designing an effective user interface. The test was a computer program in two forms; Form A and Form B. The testing platform offered item level scores, total scores, and time spent taking the test. The questions on the test were previously ranked by difficulty level from NAEP, and they were evenly distributed between the two forms. The reading comprehension test is contained in Appendix C. Each form was created for two conditions: non-accommodated and accommodated. The accommodations offered were extra time and audio presentation, as these are two of the most common accommodations offered to students with a variety of disabilities (Cormier et al., 2010; Lovett, 2010). Audio presentation was allowed for both the reading passages and the questions. A training module was incorporated into the first session of testing to allow students to practice using this new platform. This was in order to allow students to have equal opportunities to develop "test-wiseness" and reduce constraints in accessibility (Kettler et al., 2011). The training scripts are contained in Appendix D.

The testing platform had the reading passage on the left side of the screen throughout the session. Students were able to scroll up and down, and read aloud portions of the text by

paragraph. The questions were on the right side of the screen, with the option to be read aloud as well. Students clicked on a sound button positioned near the text to access audio. At the bottom of the screen a status bar with forward and backward arrows, current time, and time remaining were displayed. To access the extra time, students were presented with the option to add time when the original time concluded. During the training students were told that this option would be available. Figure 4 depicts the final design of the user interface.

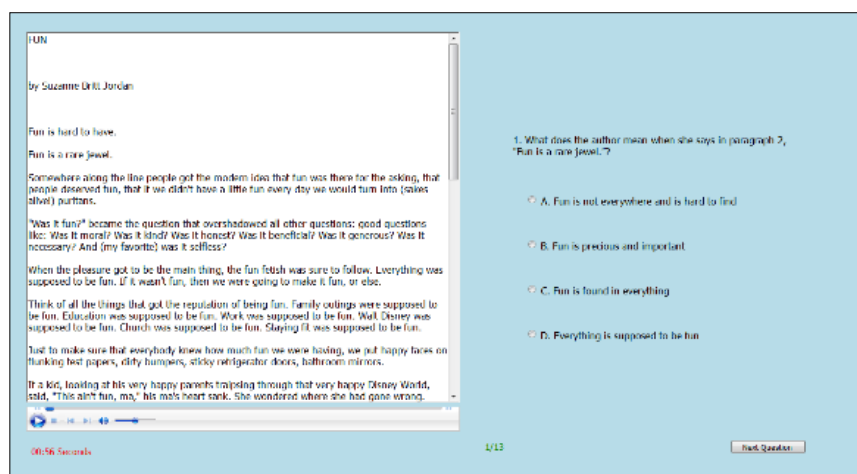


Figure 4. Screenshot of final computer program design.

Stage 2: Evaluation of Accessibility. The testing platform was evaluated for accessibility using the Test Accessibility and Modification Inventory (TAMI; Beddow et al., 2009). Four trained graduate students completed the TAMI Accessibility Rating Matrix to ensure the accessibility of the measure. Access of a measure allows for a test taker to demonstrate his/her knowledge on the tested construct (Beddow, 2012) thereby allowing for reliable scores and valid inferences.

Stage 3: Field Testing. Students were recruited from three high schools in New Jersey. The schools were initially contacted and asked if they would like to participate, and then students

were recruited from these schools. To ensure an adequate number of participants in the group with functional impairment, partnerships were made with resource room and basic skills teachers to do the assessments during their class time. The schools were offered a professional development session on the evidence based selection of accommodations for their participation. Students were offered incentives in the form of gift cards and raffles for participation. Students were then individually administered three ORF CBMs by trained graduate students. The median score from the three administrations was used to classify the students into the SWFI and SWOFI groups. The students completed two versions of the computer-based reading comprehension test (Form A and Form B). Half of the students completed Form A first, and the other half completed Form B first. Likewise, half of the students completed the accommodated condition first and the non-accommodated condition second, and the other half was tested in the opposite order. Students were randomly assigned to conditions within their group. Figure 5 depicts the design of this study. Approval to conduct this study was obtained from the Institutional Review Board at Rutgers, The State University of New Jersey prior to data collection at the schools.

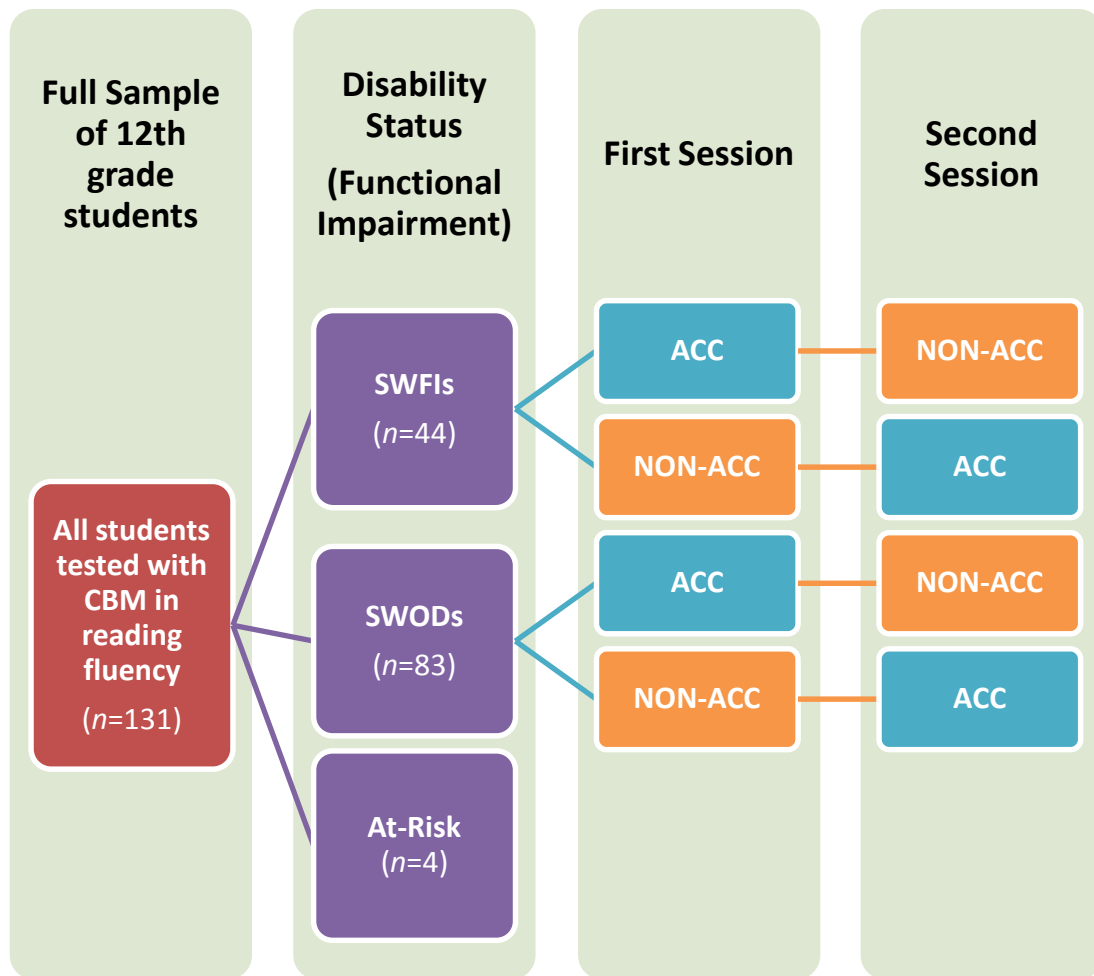


Figure 5. UD&A Study design. SWFI = students with functional impairment, SWOFI = students without functional impairment, ACC = accommodated, NON-ACC = non-accommodated.

Data Analysis

Several techniques were used to analyze the collected data. Table 4 outlines the types of evidence addressed and the analyses conducted. To assess the accessibility of the items on the test descriptive data from TAMI was used; usefulness of the accommodations was analyzed by t-test to see if there was a significant difference between the accommodations accessed by SWFIs and SWOFIs.

Descriptive statistics and qualitative data from the evaluation survey added information of perceived accessibility. The qualitative process began with open coding, an approach based on

grounded theory that inductively assigns initial thematic codes to words and phrases in the responses. According to Patton (2002), this can also be known as “first cut coding” and assists in the development of codes. In this step of the analysis, codes were created separately for each of the six open ended questions. The data was then consolidated into one set of codes to gain an overall understanding of the experience, and another coding scheme for parts of the accommodations. In this process, a hypothesis was generated to identify what the overall test experience was like, with separate findings for the test presentation, audio accommodation, and extra time accommodation. Last, the data was re-examined to confirm or disconfirm the hypothesis; Patton (2002) refers to this as the deductive part of grounded theory. Consistency and inclusiveness of the coding scheme were examined, leading to the creation of new overall positive/negative categories for each of the accommodations in order to better represent the broadness of the data.

The effect of accommodations on the reliability of the test was analyzed using Cronbach’s alpha for all four groups and conditions, with Feldt’s test for equality of the alphas (Feldt, 1980).

The validity of the inferences was measured in several ways. First, a t-test was done on the difference between scores of each group between conditions with t-tests of the main effect (group and condition). This was to test for differential boost (Fuchs & Fuchs, 1999), the standard for appropriate testing accommodations. Second, Pearson correlations were used to test for linear effects between accommodations accessed and test score increases. Lastly, Pearson correlations were studied between the CBM scores (or words read correctly per minute) and the final test scores for the students with disabilities.

Table 4

Data Analyses

Research Questions	Data Used	Data Analytic Techniques
1. Accessibility and Utility	Survey data, TAMI ratings, Frequency of accommodations accessed by group	Qualitative, T-test
2. Reliability of Scores	Item level scores by group and condition	Coefficient alpha tests for equality
3. Validity of Inferences	Test scores by group and condition	T-tests of main effects and interaction
	Number of accommodations accessed and test score increases	Pearson correlation
	Words read correctly per minute and test scores for SWDs by condition	Pearson correlation

Results**Descriptive Data**

A missing value analysis indicated that less than 7% of the data was missing. Multiple imputation methods were used for the mean comparison analyses. These procedures are less biased than traditional estimation methods for data missing not at random (Baraldi & Enders, 2010). Twenty five imputations were used; following Graham, Olchowski, and Gilreath's (2007) guidelines.

Descriptive statistics of all variables found that, with the exception of number of accommodations accessed, all variables were within acceptable limits of skewness and kurtosis. The histograms for CBM scores were used to determine how to divide the SWOFI/SWFI groups. Figure 6 depicts the histogram for ORF CBM scores in the sample. As noted in Figure 6, the CBM scores fell in two general groups, from 100-150 and from 160-250, with only 4 cases falling between these two groups, and a high number of cases immediately before and after 150-160. The mean CBM score was 168.3; therefore the 83 students with scores from 160-250 were considered scoring within measurement error of the mean or higher and comprised the SWOFI group. The 44 students whose scores were at least one half of a standard deviation below the mean in the 100-150 group comprised the SWFI group. The data of the four students between 150-160 were eliminated because they fit in neither of these categories.

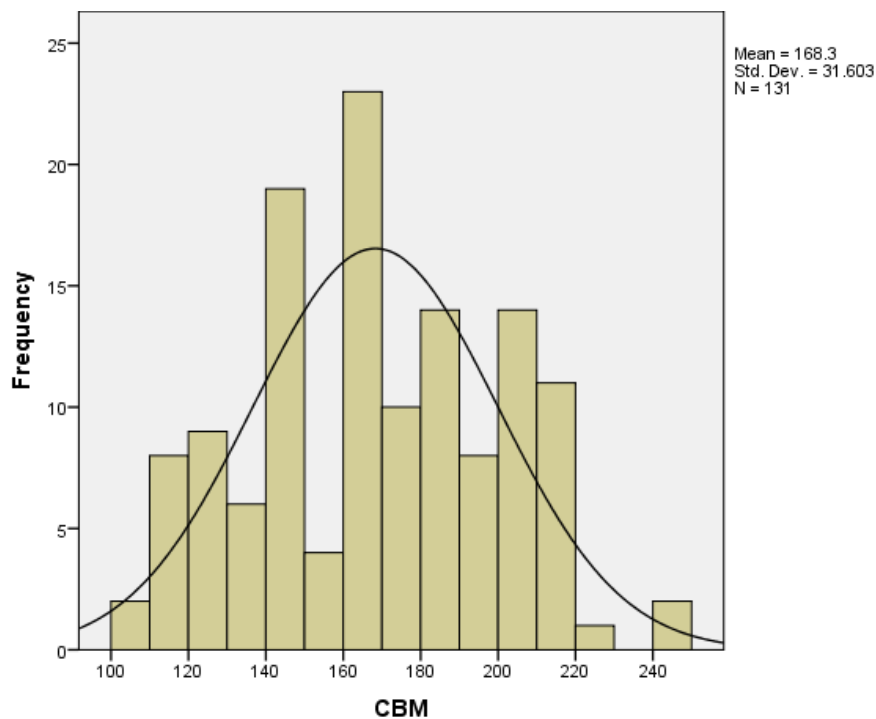


Figure 6. Frequency of CBM scores for the entire sample used for division of SWOFI and SWFI groups.

Descriptive data and exploratory t-tests indicated the SWOFI and SWFI students differed significantly in reading fluency, reading comprehension, time taken to test, and number of accommodations used. SWOFIs scored higher than SWFIs in the accommodated condition with a medium effect size ($d = .76$) and in the non-accommodated condition with a large effect size ($d = .91$). In addition, SWOFIs used less time than SWFIs in both conditions, with a medium effect size (non-accommodated $d = .63$, accommodated $d = .52$). Table 5 outlines the means and standard deviations for all variables used. Table 9 depicts the t-tests mentioned.

Table 5

Means and Standard Deviations of Scores

	SWOFI		SWFI	
	Mean	(SD)	Mean	(SD)
Measure	Oral Reading Fluency CBM			
Words Read Correctly Per Minute	187.86	(20.38)	132.66	(13.09)
	Reading Comprehension			
	(scores range from 0-26)			
Non-Accommodated Score	16.89	(5.07)	12.74	(3.97)
Accommodated Score	16.85	(5.04)	13.32	(4.16)
	Accommodations			
Number of Accommodations Accessed	2.38	(2.43)	3.01	(4.68)
(Both audio and extra time)				
Completion Time – Non-Accommodated	19.67	(7.41)	24.45	(7.84)
(50 minutes available)				
Completion Time – Accommodated	21.11	(9.12)	25.70	(8.32)
(74 minutes available)				

Note. Multiple imputation procedures were used to obtain means and standard deviations. Pooled results as obtained via SPSS were used for the means, while an average of all standard deviations was used. SWOFI = Students without functional impairment, SWFI = Students with functional impairment.

Research Question 1 Accessibility

TAMI Results. The reading comprehension tests were coded for accessibility using TAMI (Beddow et al., 2009) in the non-accommodated condition by two separate coders. Items that were not in perfect agreement were resolved by an expert coder. Of 52 items, 50 were coded in agreement, as indicated by the two separate ratings being within 1 point of each other (Beddow, Kettler, & Elliot, 2010). For 52 items, 61% were coded 3 or 4; indicating that the items were “Maximally Accessible for Nearly All/Most Test-Takers.” Another 38% of the ratings were coded 2, “Maximally Accessible for Some Test-Takers,” and no items were coded 1, “Inaccessible for Many Test-Takers.” Table 6 explains and displays the TAMI results.

Table 6

TAMI ARM Results

ARM Rating and Explanation	<i>n</i>	<i>%</i>
4 – Maximally Accessible for All Test-Takers	7	13
3 – Maximally Accessible for Most Test-Takers	25	48
2 – Maximally Accessible for Some Test-Takers	20	38
1 – Inaccessible for Many Test-Takers	0	0
Total	52	100

Note. TAMI ARM = Test Accessibility and Modification Inventory Accessibility Rating Matrix.

Item-total correlations were also considered for each item to see if accessibility impacted reliability. The item-total correlations of the non-accommodated items were positively correlated

with the item-total correlations of the accommodated items ($r = .56$). Accessibility ratings and non-accommodated item-total correlations did not evidence any relationship ($r = .18$), but evidenced a medium relationship with accommodated item-total correlations, ($r = .31$). Three items had negative item-total correlations. Two of these items were coded as 2 in accessibility, and one item was coded 3. Items coded 3 or 4 had a mean item-total correlation of .38 in the non-accommodated condition and .39 in the accommodated condition. Items coded 2 (no items were coded 1) had a mean item-total correlation of .31 in the non-accommodated condition and .26 in the accommodated condition.

Usefulness of accommodations. T-tests found that SWOFIs used significantly less accommodations than SWFIs, $t = -2.62$ (55), $p < .05$. This difference had a small effect size of $d = -.17$. The average accommodations used for group was low, with the mean SWOFI accommodations accessed 2.38 and the mean SWFI accommodations accessed 3.01. Table 9 contains the information for this comparison.

Evaluation qualitative data. One hundred twenty two students completed the evaluation study and over 160 comments in response to open-ended questions were used in the qualitative analysis. Overall, the qualitative data indicated that students who took the test found it easy. The computer presentation was viewed favorably, and the accommodations were generally viewed favorably as well. Many students wrote that they did not need accommodations, and there were some complaints about the technical and presentation aspects of the audio accommodation.

Overall test and presentation. According to the students who took the test, this was an easy test to take. Eighteen students mentioned ease of the test in their comments. For example, one student stated, “It was very easy to use. Everything was easy to read and there was no confusion [sic].” This was regardless of how well these students did on the test. One only scored

two out of thirteen correct on that section, but commented about the overall test and experience, “It was good.” However, two students commented that this test was no better or worse than regular tests. For example, one student commented, “It was generic and was no better or worse than any other computer-based reading comprehension test i [sic] have ever taken.” It is possible that by saying “easy,” students simply meant that this test fit into their regular realm of testing experiences.

Twenty nine students commented on the positive presentation of the computer based test. Specifically, students liked that the passage and questions were readily available at the same time. One student said, “The story is always on the left side and can be scrolled. We don’t have to flip any pages.” Twelve students commented on the directions and explanations accompanying the test, stating that they were helpful. One said, “The practice part was also good for people who did not understand it. That is also something that should stay in the test. It showed you how to answer questions and what everything meant [sic], so it was helpful.” Four comments disagreed with this view, with one stating, “...the tutorial part is more common sense and should not be required to take.”

Audio presentation accommodation. Although most students who used it found it helpful, there were 31 students who wrote negative comments about the technical quality and style of the audio. For instance, one wrote, “The audio was read to me in a monotone computerized voice, and it sounded staticky and harsh on the ears when I tried to listen to it using the headphones.” Another wrote, “It did not read in an efficient manner for my needs. I expected it to read faster and flow better.” Eleven students commented that they do not need the audio presentation and prefer to read on their own, such as the one who wrote, “It might be helpful for some people, but i [sic] like to read the story myself to fully understand it.” Most of

those students did not use the accommodation; some commented that it would be helpful for others.

Extra time accommodation. Twenty eight students commented that they did not need the extra time, such as one who stated, “I did not need the extra time, so that was not a problem for me.” Fifteen students wrote that the extra time was psychologically helpful to them and made them feel better about testing (none of these actually used the accommodation). One said, “It made me feel like I didn’t have to rush.” Another wrote, “It gave me time to think twice about my answers and check if they could possibly be correct or not.” These students did not do well on the test; they only received three and four correct out of 13, respectively.

As with the audio presentation accommodation, students were concerned for other students who may need it. Therefore, many students said it was helpful even if they personally did not use it. One wrote, “For the people who take longer to read its [sic] good to know they have more time to read and answer the question the best they can.”

Evaluation Quantitative Data. One hundred twenty two students completed the evaluation survey. The students responded favorably to questions regarding the overall accessibility of the test. Ninety-eight percent of respondents endorsed that the test was presented in an organized way, and 94% responded “yes” when asked if the test format was easy to use. Specific components that were incorporated in the format based on best practice review also received high favorability ratings (e.g., time remaining feature 82%, passage side by side with questions 91%, and training 67%).

Regarding the accessibility and perceived usefulness of accommodations, evaluation data was more mixed. Only 53% of respondents considering the audio presentation helpful and 61% considered extra time helpful. It is important to understand that students did not use many

accommodations altogether (mean accommodations accessed was 3.01); extra time was only used by two students in the entire sample.

Research Question 2 Reliability

Cronbach's alpha was calculated for each form by condition and then separated by group. Table 7 indicates the reliability of the forms for the entire sample and Table 8 contains the reliability information separated by group. Form A had a reliability of .83 for the non-accommodated condition and .84 for the accommodated condition. Form B had a reliability of .81 for the non-accommodated condition and .75 for the accommodated condition. An exploratory Feldt's test (Feldt & Kim, 2006) found no differences between the conditions for the overall sample in either form. Separated by group, in the SWOFI sample, alpha ranged from .79 to .86 for both forms and both conditions, which is in the acceptable range for a research test (Nunnally & Bernstein, 1994), and shows little to no change across condition. In the smaller SWFI sample, alpha was extremely variable, with Feldt's tests finding differences between forms and differences between accommodated and non-accommodated conditions. In exploring the differences between the SWOFI and SWFI alphas, Feldt's test found differences in the non-accommodated Form A; the difference was not present in the accommodated condition. For Form B, the difference was not present in the non-accommodated condition, but present in the accommodated condition. These findings are unable to be interpreted due to the small sizes of the SWFI samples. In addition, two items on each form had negative item-total correlations, which can negatively impact reliability.

Table 7

Coefficient Alpha for Reading Comprehension Test Full Sample

	Items	<i>N</i>	Crohnbach's Alpha (95% C.I.)
Non-accommodated Form A	26	61	.83 (.76-.88)
Accommodated Form A	26	62	.84 (.78-.89)
Non-accommodated Form B	26	62	.81 (.74-.87)
Accommodated Form B	26	60	.75 (.65-.83)

Note. Original data used for Coefficient Alpha analyses. SWOFI = Students without functional impairment, SWFI = Students with functional impairment.

Table 8

Coefficient Alpha for Reading Comprehension Test by Group and Form

	SWOFI			SWFI		
	Items	<i>N</i>	Crohnbach's Alpha (95% C.I.)	Items	<i>N</i>	Crohnbach's Alpha (95% C.I.)
Non-accommodated Form A	26	41	.85 (.78-.91)	26	18	.53 (.14-.79)
Accommodated Form A	26	37	.86 (.78-.92)	26	23	.80 (.67-.90)
Non-accommodated Form B	26	39	.82 (.73-.89)	26	21	.75 (.56-.88)
Accommodated Form B	26	38	.79 (.68-.87)	26	20	.36 (-.12-.70)

Note. Original data used for Coefficient Alpha analyses. SWOFI = Students without functional impairment, SWFI = Students with functional impairment.

Research Question 3 Validity

A t-test comparing the boost of scores for SWOFIs to the boost of scores for SWFIs was nonsignificant, $t = -.782$ (90), $p = .44$. No significant differential boost was found in the population. A power analysis found that with a sample size of 44 SWFIs, there was enough power to detect a large effect, but results were inconclusive for a small or medium effect as there was not enough power to detect. Table 9 depicts the mean comparisons.

Further analyses indicated that neither group's scores significantly increased from the accommodations, causing the boost variable to be similarly low in both conditions ($t = .068$, $t = -1.05$, $p > .05$). In addition, no significant relationship was found between boost and accommodations ($r = .143$, $p = .06$), possibly because there was no boost, and a low number of accommodations was used.

Table 9

Mean Comparisons

	Group		<i>t</i>	df	Cohen's d
	SWOFI	SWFI			
Student Use of Accommodations	2.38	3.01	-2.62*	55	-.17
Non-Accommodated Scores	16.89	12.74	4.53*	117	.91
Accommodated Scores	16.85	13.32	3.90*	117	.76
Difference	-0.04	0.58	-0.782	90	-.16

Note. * significant at .05, two-tailed; SWOFI = Students without functional impairment, SWFI = Students with functional impairment.

There was a small relationship between reading fluency and reading comprehension for the SWFIs in the non-accommodated condition $r = .26$, $p = .048$. This relationship lessened and exploratory analyses found that this was nonsignificant for SWFIs when accommodations were offered, $r = .12$, $p = .220$. However, an exploratory test of this difference was nonsignificant, z -score = 1.07, $p = 1.28$. Exploratory analyses indicated that the relationship stayed steady for SWOFIs regardless of condition and was similar to the relationship between reading fluency and reading comprehension for the SWFIs in the non-accommodated condition. Table 10 depicts the comparison in correlations.

Further exploratory analyses found that across the whole sample, non-accommodated and accommodated scores were highly positively correlated, CBM scores had a medium positive correlation with non-accommodated and accommodated scores, and CBM scores had a medium negative correlation with number of accommodations used. Table 11 depicts the exploratory correlations.

Table 10

Correlations Between Reading Fluency and Reading Comprehension

	SWOFI	SWFI
Non-Accommodated Scores	.25	.26*
Accommodated Scores	.22	.12

Note. * significant at .05, one-tailed; SWOFI = Students without

functional impairment, SWFI = Students with functional impairment.

Table 11

Correlations

	NON-ACC	ACC	ORF CBM	ACC USED	DIFFERENCE
NON-ACC	1	.69	.43	-.24	-.42
ACC		1	.36	-.14	.37
ORF CBM			1	-.36	-.10
ACC USED				1	.14
DIFFERENCE					1

Note. Original data (non-imputed) was used for correlational analyses. NON-ACC = non-accommodated scores, ACC = accommodated scores, ORF CBM = Oral Reading Fluency Curriculum Based Measure, ACC USED = number of accommodations used, difference is the difference between non-accommodated and accommodated scores.

Discussion

The purpose of the UD&A Study is to expand the existing research base on testing accommodations by studying the accessibility, reliability, and validity of a computer based test with universally designed accommodations. TAMI results indicated that the items were adequately accessible. Analysis of student comments and evaluations indicated a highly accessible testing platform. Regarding accommodations, students commented that they liked having accommodations available, and some complaints were noted about the audio presentation accommodation. Not many accommodations were used, possibly due to the high accessibility of the test and targeted nature of accommodations, or due to the low favorability of the audio presentation accommodation; differences were found in use of accommodations by group.

Evidence from reliability analyses was different for the different forms and underscored the importance of measuring reliability when considering accommodations as an improvement of test precision. The validity results were mixed, with data analyses indicating no increase in scores for either group with accommodations, no differential boost, and no relationship between boost and accommodations. SWFIs spent more time and used more accommodations than SWOFIs. This indicates that directly measuring access skills can predict student need and use of accommodations. The relationship between reading fluency and reading comprehension lessened for SWFIs with accommodations, indicating that the accommodations may have assisted in removing construct-irrelevant skill.

The UD&A Study builds on an extensive research base on the validity of testing accommodations as measured by differential boost. Existing studies have found mixed results for most accommodations in this arena, with some studies finding a boost and others finding no boost (Fuchs et al., 2005; Sireci et al., 2005; Cormier et al., 2010; Rogers et al., 2012). In a prior research study focusing on the audio presentation accommodation on a reading comprehension test, no boost was found due to researcher hypothesized lack of student choice in pace and amount of accommodations (Meloy et al., 2002). In another study on the extended time accommodation, no boost was found due to the hypothesized speededness of the test (Lewandowski et al., 2008). This study attempted to further delineate the appropriateness of accommodations by focusing on accessibility and reliability in addition to the traditional measure of differential boost to evaluate validity. This was done by targeting accommodations to specific functional impairments, and by analyzing the accessibility and reliability of the test. Following the recommendations of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) and the few research studies that have viewed accessibility and reliability as

necessary precursors to validity (Kettler et al., 2011), the UD&A Study has found that the assumption of appropriate accommodations having a differential boost in scores may be dependent on these precursors.

The sample included 131 twelfth grade students from both public and private high schools in New Jersey. The sample had more female than male participants and reflected the ethnic and racial diversity of New Jersey. Divided by their reading fluency CBM scores, 63% ($n = 83$) of the sample scored in the not functionally impaired range in reading fluency, and 34% ($n = 44$) of the sample scored in the functionally impaired range.

The SWOFI group obtained significantly higher scores than the SWFI group on the reading comprehension test with a large effect in the non-accommodated condition and a medium effect in the accommodated condition. The SWOFI group used significantly less time in both conditions with a medium effect, and both groups on average used much less time than was available. The SWOFI group used significantly less accommodations than the SWFI group did. Both groups used few accommodations with an average of three accommodations; the extra time accommodation was only used twice in the entire sample. The testing accommodations did not significantly raise scores for either group; the scores generally stayed flat from the non-accommodated condition to the accommodated condition.

Accessibility

As rated using the TAMI, (Beddow et al., 2009), 61% of items on the test were considered maximally accessible to most or all test-takers. This is comparable to accessibility reviews done by Beddow, Kettler, and Elliot (2013), indicating adequate accessibility of the test. In a review of items from the standardized assessments of four states Beddow et al. (2013) found that across subjects 50 to 78% of items were rated maximally accessible to most or all test-takers

with 68% of Language Arts items receiving those high ratings as well. The mean overall TAMI rating found in the UD&A Study, 2.8, was comparable to the mean overall rating found by Beddow's team.

When students rated the accessibility of the test, it garnered extremely favorable reviews, with between 80 and 98% of students endorsing the clarity, organization, and presentation style of the test. When asked about the usefulness of the accommodations, students were favorable as well, with 53 to 61% of students stating that the accommodations were useful.

The qualitative analysis of the open-ended evaluative questions dovetailed with these findings. Figure 7 displays the main qualitative findings of the study. As indicated in Figure 7, students found the test to be easy due to their positive experience with the accessible computer presentation and the psychological helpfulness of the accommodations. It is interesting to note that many of the students who used the words "easy" about the test did not in fact receive high scores, leading to the conclusion that the accessible design caused them to view the test in a positive light. This concurs with Roach et al.'s (2010) findings that students may find testing adaptations psychologically helpful even with no score benefit. Another interesting point was students' positive view of the simultaneous questions and passages on separate sides of the screen; prior research has emphasized the need to embed questions in passages (Beddow et al., 2011).

Regarding the accommodations, many students stated that they did not need them, but found them psychologically helpful, which may be important for reducing cognitive anxiety and thereby increasing test scores (Cassady & Johnson, 2002). Some technological issues and annoyance with the voice were noted for the audio presentation as well. The e-reader had a voice and cadence that is commonly used for computer-based audio presentation accommodations.

Student perception of a highly accessible test in the non-accommodated condition may have placed a cap on the ability to find different results in the accommodated condition. Alternatively, the complaints regarding the audio presentation may have held back students from using the accommodation. As indicated by these initial analyses, researcher measurement of the test found it to be adequately accessible, students found the testing to be a positive experience and the test to be highly accessible, and favorably reviewed accommodations were not used much; although used in differing degrees by the different groups.

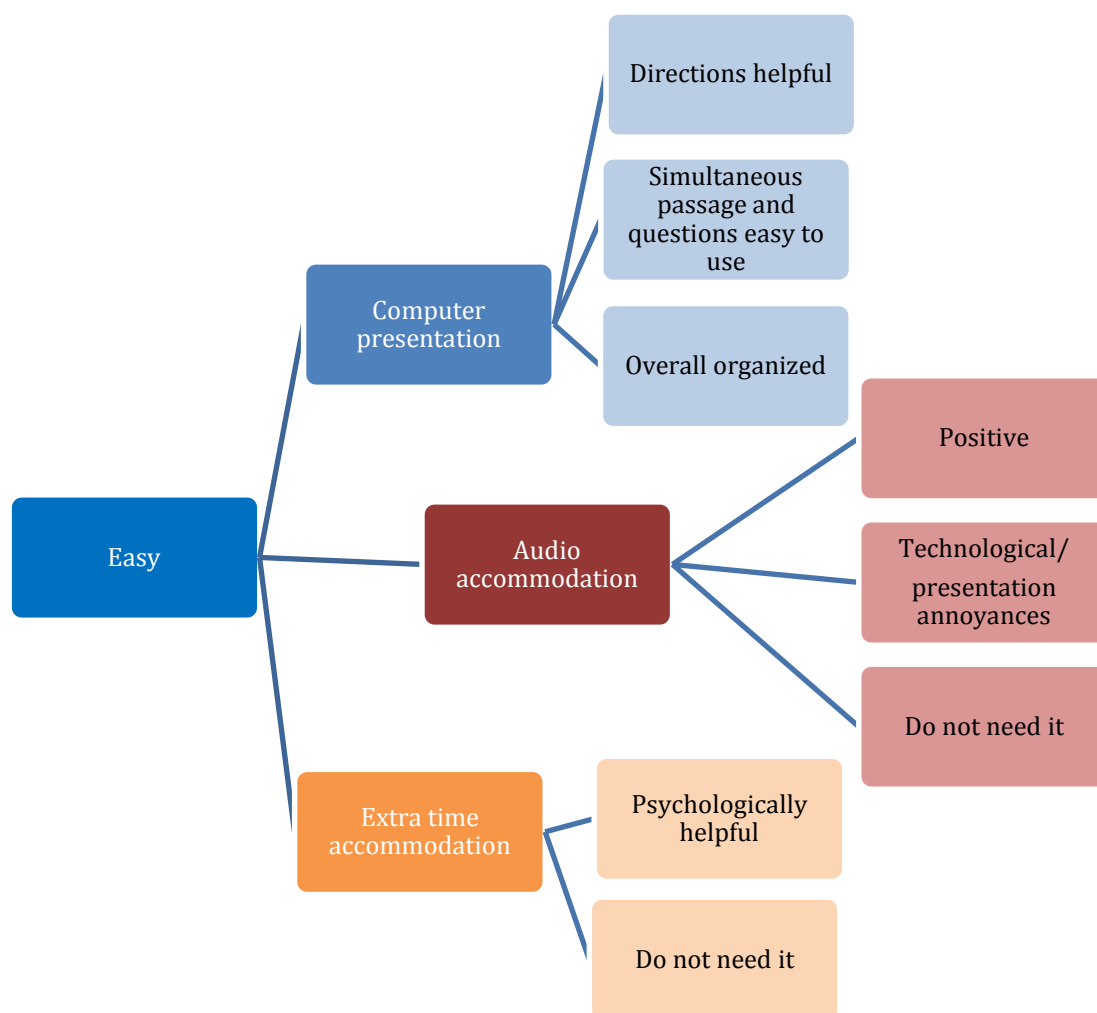


Figure 7. Major findings from UD&A qualitative analysis.

Reliability

When the SWOFI and SWFI groups were combined, both forms of the test had acceptable reliability (ranging from .75-.84) for a research test, with no differences found among the overall sample. When Cronbach's alpha was calculated for each group separately, more differences appeared. Since there were two forms, the sample sizes for the SWFI groups were very small, ranging from 18-23. This caused extreme variability in the scores and the results should be taken with caution. Again, Form A appeared to have functioned as expected, with Feldt's tests finding significant differences between groups in the non-accommodated condition and no difference in reliability in the accommodated condition. Form B did not function as expected. There were no significant differences between groups in the non-accommodated condition and there was a difference in the accommodated condition. This may be due to the differences in passages used for each form.

There has not been much research addressing the differences in reliability for tests taken by students with and without disabilities and how and whether reliability should improve with testing accommodations. When studying item modifications, Kettler et al. (2011), found that reliability generally stayed the same. Although the results from the UD&A Study need to be interpreted with caution, they underscore the need to study alpha separately by group when evaluating the precision of a measurement tool. SWFI's scores were less internally consistent than SWOFI's in almost all cases, and the accommodations were not necessarily sufficient to overcome the lesser reliability. Less reliable scores indicate that the scores are less precise and may not reflect a consistent construct (Thompson, 2002).

In this study, having such low alpha levels limit the ability to draw definitive conclusions for all of the research questions, as there may have been other confounds that contributed to

these low levels, such as executive functioning or working memory deficits. Further research is needed to ensure that additional barriers, aside from those addressed with accommodations, do not exist for SWFIs. In studying the research base and the theoretical guidelines, more definitive expectations and criteria must be set for the expected improvement of reliability with testing adaptations.

Validity

Although a differential boost was hypothesized for this study, some prior research studies have not found differential boosts for these accommodations (e.g. Meloy et al., 2002; Lewandowski et al., 2008). It was hypothesized that ensuring that testing accommodations directly target functional impairments in access skills would lead to differential boost. In the UD&A Study, accessibility was measured and quantified, and constructs were precise. Yet, neither group's scores increased significantly from accommodation. These accommodations have been controversial, and audio presentation has been categorized by some as inappropriate on a reading test due to its perceived impact on the construct knowledge (Meloy et al., 2002). Therefore, the lack of any boost in this study may raise questions about the effectiveness of these accommodations.

Although the SWOFIs used significantly less accommodations than SWFIs, not many accommodations were used by either group (averaging two or three). Due to the care taken to ensure the overall accessibility of the test, this is not surprising. Although the test was timed, there was adequate time available for students to answer all the questions and speed of answering questions was not a part of the targeted construct. Average completion times ranged from 20 to 26 minutes; in the non-accommodated version 50 minutes were available. In addition, the test's constructs were clearly targeted as critical thinking with no benefit to be gained from audio

presentation other than overcoming a functional impairment. Possibly due to this, neither the SWOFI group nor the SWFI group's scores were significantly raised from the accommodations. With no boost in either group, no differential boost was found either.

The descriptive results provide justification for the paradigm used in offering accommodations. Only three students identified as Special Education students and 44 were functionally impaired in reading fluency. Standardized measurement of this access skill was able to predict longer time taken on the test and more accommodations used, indicating that deficits in reading fluency require students to use more time and more accommodations. A lessened relationship between reading fluency and scores for the SWFI group in the accommodated condition can also provide a justification for offering these accommodations.

As outlined in Figure 2, in order for valid inferences to be drawn from accommodated test scores, individual need should be considered, and the construct should be unchanged, so that differential effects are found, and inferences can be considered valid. In the UD&A Study, accommodations were directly tied to individual student need for assistance in overcoming the access barrier of reading fluency. The critical thinking construct on the test was unchanged. The third step of differential effects ensures that accommodations should not assist SWOFIs as much as SWFIs, because that may give an unfair advantage to SWFIs. In this study, no differential boost was found. Much of the research base is concerned regarding the possibility that accommodations raise scores for all students and therefore provide an unfair advantage to students who receive them (Fuchs & Fuchs, 2001). In this study no boosts were found at all. It is plausible that all four groups of scores (SWOFI non-accommodated and accommodated; SWFI non-accommodated and accommodated) can have sameness of inferences for this test. The

student-perceived high accessibility and targeted construct may have obviated the need for accommodations, with the Universal Design functioning as intended.

Summary of Findings

Considered as a whole, the findings of this study support the promise of using UD to provide accommodations as a choice for all students. The accessibility evidence underscores the need for consideration of all student needs as building blocks for any test, especially through targeting the constructs and ensuring accessible system and test design. The accessibility of the test in this study may have limited the use and benefit of accommodations. Alternatively, a low use of accommodations may have been due to the unfavorably rated computer voice and the non-speeded nature of the test.

Additionally, this study was an important step toward understanding how accommodations, when properly implemented, can increase the measurement precision of a test. Although inconclusive, the reliability evidence showed that reliability can be significantly less for SWFIs without accommodations. Also, qualitative data from students indicated the value of accommodations in decreasing testing anxiety.

The traditional measure of validity, differential boost, did not occur as neither group of students gained significantly from the accommodations. However, SWFIs used significantly more accommodations and more time than SWOFIs, suggesting that the accommodations were accessed and used more by those for whom they were intended. Correlations between reading fluency and reading comprehension also provided evidence for the accommodations targeting reading fluency. No evidence from this study found that the validity of the test was compromised with testing accommodations. The accommodations of audio presentation and extra time appeared to function as an access ramp accommodation would in a test of crossing the street -

used more by those who need it and giving no edge - as they did not alter the construct of critical thinking.

Limitations

One limitation of this study was that the students were voluntarily taking a test for research purposes and knew that their scores would not be used for grades or decision making. Therefore, students perhaps did not spend as much time or effort as they may have had the test been a true high-stakes test.

A second limitation in the study is the small sample of the SWFIs. Since this was a voluntary test and a convenience sample, less students with difficulty reading were available or elected to participate. This limited the power of the reliability and differential boost analyses.

A third limitation in the study is that only twelfth grade students were sampled. For high school, the decision to separate the construct of reading fluency from critical thinking was plausible. In younger grades, these two constructs may be too closely tied for the audio presentation accommodation to be valid.

Future Research

More research should be done to further explore whether accessibility may be more important for the validity of scores than testing accommodations. Specifically, studies can compare tests with extensive accessibility review to tests with accommodations to see whether reliability and validity are impacted by one more than the other. Future research should also replicate the findings of the study that there was no loss of time to give accommodations universally. With regard to accommodations that appeared to give an unfair edge to some students, researchers have spent years answering the question of why students with disabilities

should have accommodations. Having accommodations targeted specifically to access skills begs a new question of why these accommodations should not be universally included in test design.

Additionally, future research should focus on reliability in order to answer key questions about whether testing adaptations are expected to improve reliability and by how much. Lastly, future research should continue to integrate student perspectives. With a focus on accessibility, students must be able to voice their perception of testing features and quality.

Conclusions

IDEA (1997) and IDEIA (2004) require states to provide equitable assessment for students with disabilities, in order to ensure that they will be subject to high expectations and educational opportunity (Fuchs, Fuchs, & Capizzi, 2005; Kettler & Elliott, 2010). The principles of UD (Center for Universal Design, 1997; Story, Mueller, & Mace, 1998) and the new focus on accessibility (Beddow et al., 2011) provide a powerful way to create inclusive assessments. Prior research has mostly focused on the differential boost of testing accommodations for students with disabilities and has found mixed results. The UD&A Study used accessibility review and principles to create a reading comprehension test with audio presentation and extra time accommodations. The evidence from this study indicates that using the principles of UD to create highly accessible tests may have superseded the need for many accommodations. In addition, accommodations were targeted to functional impairments in the access skill of reading fluency. This ensured that SWOFIs used less accommodations, and that SWFI scores on reading comprehension tests were a valid representation of critical thinking.

Although reliability evidence was mixed, there is promise that reliability can improve from targeted accommodations. Neither group of students' scores increased significantly or differentially. However, the goal of accommodations is not to facilitate higher scores; it is to

provide access to the test content and reliable scores in order that valid inferences can be drawn.

The UD&A Study provided a promising direction for exploring accessibility and validity evidence that supports this case.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Author.
- Bader, L. A. (2005) *Bader reading and language inventory, 5th edition*. Pearson.
- Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*(1), 5-37.
- Beddow, P. A. (2012). Accessibility theory for enhancing the validity of test results for students with special needs. *International Journal of Disability, Development and Education, 59*(1), 97-111. doi: 10.1080/1034912X.2012.654966
- Beddow, P. A., Elliott, S. N., & Kettler, R. J. student (2009). *TAMI accessibility rating matrix (ARM)*. Nashville, TN: Vanderbilt University.
- Beddow, P. A., Elliot, S. N., & Kettler, R. J. (2010). *TAMI accessibility rating matrix technical manual*. Nashville, TN: Vanderbilt University.
- Beddow, P. A., Elliot, S. N., & Kettler, R. J. (2013). Test accessibility: Item reviews and lessons learned from four state assessment. *Educational Research International, 2013*, Article ID 952704. doi: 10.1155/2013/952704
- Beddow, P.A., Kurz, A., & Frey, J.R. (2011). Accessibility theory: Guiding the science and practice of test item design with the test-taker in mind. In S.N. Elliott et al. (Eds.) *Handbook of Accessible Achievement Tests for All Students*, (pp. 163-182). Springer.
- Bolt, S. E., & Thurlow, M. L. (2004). Five of the most frequently allowed testing accommodations in state policy synthesis of research. *Remedial and Special Education, 25*(3), 141-152. doi: 10.1177/07419325040250030201

- Brown, J. I., Fishco, V. V., & Hanna, G. (1993). *Nelson-Denny Reading Test, Form H*. Itasca, IL: Riverside Publishing.
- Carter, E.W., Wehby, J., & Hughes, C. (2005). Preparing adolescents with high-incidence disabilities for high-stakes testing with strategy instruction. *Preventing School Failure*, 49(2), 55-62. doi: 10.1080/1045988X.2005.10823218
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27(2), 270-295.
- Center for Universal Design. (1997). *What is universal design?* NC State University: The Center for Universal Design.
- 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.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Cormier, D.C., Altman, J.R., Shyyan, V., & Thurlow, M.L. (2010). *A summary of the research on the effects of test accommodations: 2007-2008*. (Technical Report 56). Minneapolis, MN: University of Minnesota: National Center on Educational Outcomes.
- Csapo, B., Ainley, J., Bennett, R. E., Latour, T., & Law, N. (2012). Technological issues for computer-based assessment. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 143-230). Dordrecht, Netherlands: Springer.
- CTB/McGraw-Hill. (2000). *Guidelines for using the results of standardized tests administered under nonstandard conditions*. Monterey, CA: Author.
- CTB/McGraw-Hill. (2005). *Guidelines for Inclusive Test Administration*. Monterey, CA: Author.

- Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of UD 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). Available from <http://www.jtla.org>
- Educational Testing Service (2013). *Assessing people with disabilities*. Retrieved from http://www.ets.org/research/topics/assessing_people_with_disabilities
- Elliott, S. N., Kettler, R. J., Beddow, P. A., Kurz, A., Compton, E., McGrath, D., ... & Roach, A. T. (2010). Effects of using modified items to test students with persistent academic difficulties. *Exceptional Children*, 76(4), 475-495.
- Elliott, S. N., Kratochwill, T. R., & McKevitt, B. C. (2001). Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities. *Journal of School Psychology*, 39(1), 3-24. doi: 10.1016/S0022-4405(00)00056-X
- Feldt, L. S., & Kim, S. (2006). Testing the difference between two alpha coefficients with small samples of subjects and raters. *Educational and Psychological Measurement*, 66(4), 589-600.
- Fuchs, L. S., & Fuchs, D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research and Practice*, 16, 174–181. doi: 10.1111/0938-8982.00018
- Fuchs, L. S., Fuchs, D., & Capizzi, A. M. (2005). Identifying appropriate test accommodations for students with learning disabilities. *Focus on Exceptional Children*, 37(6), 1-8.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L. & Germann, G. (1993). Formative evaluation of academic progress: How much growth should we expect? *School Psychology Review*, 22, 27-48.

- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3), 239-256. doi: 10.1207/S1532799XSSR0503_3
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8(3), 206-213.
- Higgins, J., Fedorchak, G., & Katz, M. (2012). *Assignment of Accessibility Tools for Digitally Delivered Assessments: Key Findings*. Dover, NH: Measured Progress.
- Hollenbeck, K. (2005). Validity issues and decisions about testing accommodations. *Assessment for Effective Intervention*, 31(7), 7-17. doi: 10.1177/073724770503100102
- 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: validity, technical adequacy, and implementation*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Hopkins, W. C. (2001). A scale of magnitudes for effect statistics. *A New View of Statistics*. Retrieved from <http://www.sportsci.org/resource/stats/effectmag.html>
- Hosp, M. K., Hosp, J. L., & Howell, K. W. (2012). *The ABCs of CBM: A practical guide to curriculum-based measurement*. Guilford Press.
- Individuals with Disabilities Education Act of 1997. Pub. L., 101-476, 104 Stat. 1142.
- Individuals with Disabilities Education Improvement Act 2004. Pub. L., 208-446, 118 Stat. 2647.
- Kaplan, R.M. (1987). *Basic statistics for the behavioral sciences*. Boston, MA: Allyn and Bacon.

- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*(3), 281-300. doi: 10.1080/10888430802132279
- Ketterlin-Geller, L. R. (2005). Knowing what all students know: Procedures for developing UD for assessment. *Journal of Technology, Learning, and Assessment, 4*(2). Available from <http://www.jtla.org>
- Kettler, R. J. (2011). Holding modified assessments accountable: Applying a unified reliability and validity framework to the development and evaluation of AA-MASs. In M. Russell & M. Kavanaugh (Eds.), *Assessing Students in the Margins: Challenges, Strategies, and Techniques* (pp. 311–334). Charlotte, NC: Information Age Publishing.
- Kettler, R. J. (2012). Testing Accommodations: Theory and research to inform practice. *International Journal of Disability, Development and Education, 59*(1), 53-66. doi: 10.1080/1034912X.2012.654952
- Kettler, R.J., Braden, J.P., & Beddow, P.A. (2011). Test-taking skills and their impact on accessibility for all students. In S.N. Elliott et al. (Eds.) *Handbook of Accessible Achievement Tests for All Students*, (pp. 147-159). Springer.
- Kettler, R. J., & Elliott, S. N. (2010). Assessment accommodations for children with special needs. In E. Baker, P. Peterson, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., pp. 530–536). Oxford: Elsevier. doi: 10.1016/B978-0-08-044894-7.01130-1
- Kettler, R. J., & Feeney-Kettler, K. A. (2011). Screening systems and decision making at the preschool level: Application of a comprehensive validity framework. *Psychology in the Schools, 48*(5), 430-441. doi: 10.1002/pits.20565

- Kettler, R.J., Feeney-Kettler, K.A., Palladino, M.A., Zahra, L.P., & Rodriguez, J.C. (2013). A comprehensive framework for evaluating systems for screening preschool behaviors. In C.H. Qi & T. Stanton-Chapman (Eds.). *Preschool children: Education, language, social functioning and behavioral issues*. Hauppauge, New York: Nova Science Publishers, Inc.
- Kettler, R. J., Rodriguez, M. R., Bolt, D. M., Elliott, S. N., Beddow, P. A., & Kurz, A. (2011). Modified multiple-choice items for alternate assessments: Reliability, difficulty, and differential boost. *Applied Measurement in Education*, 24 (3), 210-234. doi: 10.1080/08957347.2011.580620
- King, J. E. (2011). Implementing the Common Core state standards: An action agenda for higher education. Retrieved from <http://www.acenet.edu/news-room/Documents/Implementing-the-Common-Core-State-Standards-2011.pdf>
- Laitusis, C. C. (2010). Examining the impact of audio presentation on tests of reading comprehension. *Applied Measurement in Education*, 23, 153-167. doi: 10.1080/08957341003673815
- Lewandowski, L. J., Lovett, B. J., & Rogers, C. L. (2008). Extended time as a testing accommodation for students with reading disabilities does a rising tide lift all ships? *Journal of Psychoeducational Assessment*, 26(4), 315-324. doi: 10.1177/0734282908315757
- Lovett, B. J. (2010). Extended time testing accommodations for students with disabilities answers to five fundamental questions. *Review of Educational Research*, 80(4), 611-638. doi: 10.3102/0034654310364063

- 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*(4), 248-255. doi: 10.1177/07419325020230040801
- National Assessment of Educational Progress (2013). *Reading Framework for the 2013 National Assessment of Educational Progress*. Retrieved from:
<http://nces.ed.gov/nationsreportcard/reading/moreabout.aspx>
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory (3rd edition)*. McGraw-Hill, Inc., New York: NY, 264-265.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods, 3rd edition*. Thousand Oaks, CA: Sage.
- Partnership for Assessment of Readiness for College and Careers (2013). *Accessibility features and accommodations manual, 1st edition*. Retrieved from:
<http://www.parcconline.org/parcc-assessment-policies>
- Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices—past, present, and future. In S. G. Paris & S. A. Stahl (Eds.), *Children's Reading Comprehension and Assessment* (pp. 13–69). Mahwah, NJ: Erlbaum.
- Pinnell, G.S., Pikulski, J.J., Wixson, K.K., Campbell, J.R., Gough, P.B., & Beatty, A.S. (1995). *Listening to children read aloud*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Rasinski, T. V., Padak, N. D., McKeon, C. A., Wilfong, L. G., Friedauer, J. A., & Heim, P. (2005). Is reading fluency a key for successful high school reading? *Journal of Adolescent & Adult Literacy, 49*(1), 22-27. doi: 10.1598/JAAL.49.1.3

- Roach, A. T. & Beddow, P. A. (2011). Including student voices in the design of more inclusive assessments. In S.N. Elliott et al. (Eds.) *Handbook of Accessible Achievement Tests for All Students*, (pp. 243-254). Springer.
- Roach, A. T., Beddow, P. A., Kurz, A., Kettler, R. J., & Elliott, S. N. (2010). Incorporating student input in developing alternate assessments based on modified academic achievement standards. *Exceptional Children*, 77, 61-80.
- Rogers, C. M., Christian, E. M., & Thurlow, M. L. (2012). *A summary of the research on the effects of test accommodations: 2009-2010* (Technical Report 65). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Russell, M., Hoffman, T., & Higgins, J. (2009). NimbleTools: A universally designed test delivery system. *Teaching Exceptional Children*, 42(2), pp 6-12.
- Sarroub, L., & Pearson, P. D. (1998). Two steps forward, three steps back: The stormy history of reading comprehension assessment. *The Clearing House*, 72 (2), 97-105. doi: 10.1080/00098659809599604
- Shneiderman, B. (2003). *Designing the user interface*. Pearson Education India.
- Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, 21(3), 469-479.
- Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75(4), 457–490. doi: 10.3102/00346543075004457

- Smarter Balanced Assessment Consortium (2013). *Usability, accessibility, and accommodations guidelines*. Retrieved from: http://www.smarterbalanced.org/wordpress/wp-content/uploads/2013/09/SmarterBalanced_Guidelines_091113.pdf
- Story, M. F., Mueller, J. L., & Mace, R. L. (1998). *The universal design file: Designing for people of all ages and abilities*. NC State University: The Center for Universal Design.
- Thompson, B. (2002). *Score reliability: Contemporary thinking on reliability issues*. Sage publications.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *UD applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved September 30, 2013, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>
- Thurlow, M., House, A., Boys, C., Scott, D., & Ysseldyke, J. (2000). *State participation and accommodation policies for students with disabilities: 1999 Update* (Synthesis Rep. No. 33). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Appendix A

Demographic Questionnaire

Please answer the following questions:

1. I am _____ years old
2. I am a
 - a. Female
 - b. Male
3. My ethnicity is
 - a. White
 - b. Black
 - c. Asian/Pacific Islander
 - d. Latino/Latina
 - e. Native American
 - f. Other
4. I am currently receiving Special Education services
 - a. Yes
 - b. No

Appendix B

Evaluation Survey

The following questions address your experience in taking the computer reading comprehension test:

1. Have you ever taken a reading comprehension test on a computer before participating in this study?
 - a. Yes
 - b. No
2. Do you think this test was presented in an organized way?
 - a. Yes
 - b. No

Please explain

3. Did you find the test format easy to use?
 - a. Yes
 - b. No

Please explain

4. Did you find the time remaining feature (pictured) helpful?
 - a. Yes
 - b. No
5. Did you find the passage on the side helpful?
 - a. Yes
 - b. No
6. Did you find the training module at the beginning of the testing session helpful?
 - a. Yes
 - b. No
7. Did you find the audio presentation of the passages and questions helpful?
 - a. Yes
 - b. No

Please explain

8. Did you find the extra time option helpful?
 - a. Yes
 - b. No

Please explain

9. Was it easy to understand how to access the audio presentation and the extra time?
 - a. Yes
 - b. No
 - c. If no, what could have helped?
10. Please share any further information about your experience taking this test that could be helpful. Thank you!

Appendix C

Reading Comprehension Test

Passages 1 and 2: (Form A accommodated, Form B non-accommodated)

Reading Passage [1]

FUN

by Suzanne Britt Jordan

Fun is hard to have.

Fun is a rare jewel.

Somewhere along the line people got the modern idea that fun was there for the asking, that people deserved fun, that if we didn't have a little fun every day we would turn into (sakes alive!) puritans.

"Was it fun?" became the question that overshadowed all other questions: good questions like: Was it moral? Was it kind? Was it honest? Was it beneficial? Was it generous? Was it necessary? And (my favorite) was it selfless?

When the pleasure got to be the main thing, the fun fetish was sure to follow. Everything was supposed to be fun. If it wasn't fun, then we were going to make it fun, or else.

Think of all the things that got the reputation of being fun. Family outings were supposed to be fun. Education was supposed to be fun. Work was supposed to be fun. Walt Disney was supposed to be fun. Church was supposed to be fun. Staying fit was supposed to be fun.

Just to make sure that everybody knew how much fun we were having, we put happy faces on flunking test papers, dirty bumpers, sticky refrigerator doors, bathroom mirrors.

If a kid, looking at his very happy parents traipsing through that very happy Disney World, said, "This ain't fun, ma," his ma's heart sank. She wondered where she had gone wrong. Everybody told her what fun family outings to Disney World would be. Golly gee, what was the matter?

Fun got to be such a big thing that everybody started to look for more and more thrilling ways to supply it. One way was to step up the level of danger so that you could be sure that, no matter what, you would manage to have a little fun.

Television commercials brought a lot of fun and fun-loving folks into the picture. Everything that people in those commercials did looked like fun: taking Polaroid snapshots, buying insurance, mopping the floor, bowling, taking aspirin. The more commercials people watched, the more they wondered when the fun would start in their own lives. It was pretty depressing.

Big occasions were supposed to be fun. Christmas, Thanksgiving and Easter were obviously supposed to be fun. Your wedding day was supposed to be fun. Your honeymoon was supposed to be the epitome of fun. And so we ended up going through every Big Event we ever celebrated, waiting for the fun to start.

It occurred to me, while I was sitting around waiting for the fun to start, that not much is, and that I should tell you just in case you're worried about your fun capacity.

I don't mean to put a damper on things. I just mean we ought to treat fun reverently. It is a mystery. It cannot be caught like a virus. It cannot be trapped like an animal. The god of mirth is paying us back for all those years of thinking fun was everywhere by refusing to come to our party. I don't want to blaspheme fun anymore. When fun comes in on little dancing feet, you probably won't be expecting it. In fact, I bet it comes when you're doing your duty, your job, or your work. It may even come on a Tuesday.

I remember one day, long ago, on which I had an especially good time. Pam Davis and I walked to the College Village drug store one Saturday morning to buy some candy. We were about 12 years old. She got her Bit-O-Honey. I got my malted milk balls, chocolate stars, Chunkys, and a small bag of M & M's. We started back to her house. I was going to spend the night. We had the whole day to look forward to. We had plenty of candy. It was a long way to Pam's house but every time we got weary Pam would put her hand over her eyes, scan the horizon like a sailor and say, "Oughta reach home by nightfall," at which point the two of us would laugh until we thought we couldn't stand it another minute. Then after we got calm, she'd say it again. You should have been there. It was the kind of day and friendship and occasion that made me deeply regretful that I had to grow up.

It was fun.

From The New York Times, December 13, 1979, copyright © 1979 by The New York Times. Used by permission.

1. What does the author mean when she says in paragraph 2, "Fun is a rare jewel."?
 - A. Fun is not everywhere and is hard to find
 - B. Fun is precious and important
 - C. Fun is found in everything
 - D. Everything is supposed to be fun
2. In paragraph 3, when the author mentions the possibility of people turning into puritans, she is using this word to refer to people who
 - A. lived a long time ago
 - B. rarely make a mistake
 - C. are serious and reserved
 - D. dress in plain and dark clothing

3. At the beginning of the essay, the author suggests that people are so concerned with having fun that they
 - A. try to find fun in all their experiences
 - B. spend a lot of money trying to have fun
 - C. join groups to learn how to have fun
 - D. avoid new experiences that may not be fun
4. How does paragraph 10 about television commercials relate to the author's main point about fun?
 - A. Television commercials make people want to have fun, and fun is not so important
 - B. Television commercials make it look like fun is everywhere; while it is actually rare and special
 - C. Television commercials can make people jealous when others have fun, and that causes less fun
 - D. Television commercials can be annoying to watch because there is too much fun emphasized
5. What is the author's point about big occasions like holidays?
 - A. They go by too quickly to be enjoyed.
 - B. They are not as much fun as people expect them to be.
 - C. They have become too centered around money.
 - D. They help us to appreciate the important events in life.
6. The author ends the essay with a childhood story. Why would she employ this technique?
 - A. To show that fun really can be found everywhere
 - B. To show that a lot of fun happens to younger children
 - C. To persuade readers with a personal example
 - D. To add her own experiences of fun that were not rare
7. The author assumes that the people reading her essay
 - A. probably had fun going to amusement parks as children
 - B. prefer dangerous experiences over fun activities
 - C. may be worried that they are not having enough fun
 - D. enjoy discussing the topic of fun
8. In paragraph 13, when the author tells us to "treat fun reverently," she is encouraging us to
 - A. look forward to having fun
 - B. have great respect for fun
 - C. teach others how to have fun
 - D. have fun less frequently
9. What overall tone or voice did the author have in the essay?
 - A. A sarcastic tone; she said that "work was supposed to be fun"
 - B. A depressed tone; she does not like to have fun

- C. A happy tone; she talked about weddings, honeymoons, and Disney
- D. A reminiscing voice; she retold stories from her childhood

The following paragraph is from the essay.

Think of all the things that got the reputation of being fun. Family outings were supposed to be fun. Education was supposed to be fun. Work was supposed to be fun. Walt Disney was supposed to be fun. Church was supposed to be fun. Staying fit was supposed to be fun.

10. What is the author implying in paragraph 6?
 - A. It is possible to have fun in a wide range of activities.
 - B. A person's reputation is based on how much fun the person has.
 - C. Most daily activities are less important than we think.
 - D. We should not expect everything in life to be fun.
11. "Was it fun?" became the question that overshadowed all other questions: good questions like: Was it moral? Was it kind? Was it honest? Was it beneficial? Was it generous? Was it necessary? And (my favorite) was it selfless?
What is the main idea the author is trying to bring out with paragraph 4?
 - A. Fun has come to take the place of education, work, and church
 - B. The focus on fun has overshadowed other important considerations
 - C. Something can be fun and also be moral, kind, honest, and beneficial
 - D. Even asking the question "Was it fun?" can make you depressed
12. According to the author, what could be a problem with a constant pursuit of fun?
 - A. It can cause people to be depressed when they do not have fun
 - B. It can cause people to have unrealistic expectations
 - C. It makes people think that they should be living lives similar to those on TV
 - D. It can lead people to dangerous choices
13. According to the author, fun is most likely to come
 - A. On big and exciting occasions
 - B. When you really want it
 - C. Unexpectedly
 - D. When you try hard to make something fun

Reading Passage [2]

The Flying Machine
By Ray Bradbury

In the year A.D. 400, the Emperor Yuan held his throne by the Great Wall of China, and the land was green with rain, readying itself toward the harvest, at peace, the people in his dominion neither too happy nor too sad.

Early on the morning of the first day of the first week of the second month of the new year, the Emperor Yuan was sipping tea and fanning himself against a warm breeze when a servant ran across the scarlet and blue garden tiles, calling, "Oh, Emperor, Emperor, a miracle!"

"Yes," said the Emperor, "the air is sweet this morning."

"No, no, a miracle!" said the servant, bowing quickly.

"And this tea is good in my mouth, surely that is a miracle."

"No, no, Your Excellency."

"Let me guess then - the sun has risen and a new day is upon us. Or the sea is blue. That now is the finest of all miracles."

"Excellency, a man is flying!"

"What?" The Emperor stopped his fan.

"I saw him in the air, a man flying with wings. I heard a Voice call out of the sky, and when I looked up, there he was, a dragon in the heavens with a man in its mouth, a dragon of paper and bamboo, coloured like the sun and the grass."

"It is early," said the Emperor, "and you have just wakened from a dream."

"It is early, but I have seen what I have seen! Come, and you will see it too."

"Sit down with me here," said the Emperor. "Drink some tea. It must be a strange thing, if it is true, to see a man fly. You must have time to think of it, even as I must have time to prepare myself for the sight." They drank tea.

"Please," said the servant at last, "or he will be gone." The Emperor rose thoughtfully. "Now you may show me what you have seen." They walked into a garden, across a meadow of grass, over a small bridge, through a grove of trees, and up a tiny hill.

"There!" said the servant. The Emperor looked into the sky.

And in the sky, laughing so high that you could hardly hear him laugh, was a man; and the man was clothed in bright papers and reeds to make wings and a beautiful yellow tail, and he was soaring all about like the largest bird in a universe of birds, like a new dragon in a land of ancient dragons.

The man called down to them from high in the cool winds of morning. "I fly, I fly!"

The servant waved to him. "Yes, yes!"

The Emperor Yuan did not move. Instead he looked at the Great Wall of China now taking shape out of the farthest mist in the green hills, that splendid snake of stones which writhed with majesty across the entire land. That wonderful wall which had protected them for a timeless time from enemy hordes and preserved peace for years without number. He saw the town, nestled to itself by a river and a road and a hill, beginning to waken.

"Tell me," he said to his servant, "has anyone else seen this flying man?"

"I am the only one, Excellency," said the servant, smiling at the sky, waving.

The Emperor watched the heavens another minute and then said, "Call him down to me."

"Ho, come down, come down! The Emperor wishes to see you!" called the servant, hands cupped to his shouting mouth.

The Emperor glanced in all directions while the flying man soared down the morning wind. He saw a farmer, early in his fields, watching the sky, and he noted where the farmer stood. The flying man alit with a rustle of paper and a creak of bamboo reeds. He came proudly to the Emperor, clumsy in his rig, at last bowing before the old man.

"What have you done?" demanded the Emperor.

"I have flown in the sky, Your Excellency," replied the man.

"What have you done?" said the Emperor again.

"I have just told you!" cried the flier.

"You have told me nothing at all." The Emperor reached out a thin hand to touch the pretty paper and the birdlike keel of the apparatus. It smelled cool, of the wind.

"Is it not beautiful, Excellency?"

"Yes, too beautiful."

"It is the only one in the world!" smiled the man. "And I am the inventor."

"The only one in the world?"

"I swear it!"

"Who else knows of this?"

"No one. Not even my wife, who would think me mad with the son. She thought I was making a kite. I rose in the night and walked to the cliffs far away. And when the morning breezes blew and the sun rose, I gathered my courage, Excellency, and leaped from the cliff. I flew! But my wife does not know of it."

"Well for her, then," said the Emperor. "Come along."

They walked back to the great house. The sun was full in the sky now, and the smell of the grass was refreshing.

The Emperor, the servant, and the flier paused within the huge garden. The Emperor clapped his hands. "Ho, guards!" The guards came running. "Hold this man." The guards seized the flier. "Call the executioner," said the Emperor. "What's this!" cried the flier, bewildered. "What have I done?" He began to weep, so that the beautiful paper apparatus rustled.

"Here is the man who has made a certain machine," said the Emperor, "and yet asks us what he has created. He does not know himself. It is only necessary that he create, without knowing why he has done so, or what this thing will do." The executioner came running with a sharp silver ax. He stood with his naked, large-muscled arms ready, his face covered with a serene white mask.

"One moment," said the Emperor. He turned to a nearby table upon which sat a machine that he himself had created. The Emperor took a tiny golden key from his own neck. He fitted his key to the tiny, delicate machine and wound it up. Then he set the machine going. The machine was a garden of metal and jewels. Set in motion, the birds sang in tiny metal trees, wolves walked through miniature forests, and tiny people ran in

and out of sun and shadow, fanning themselves with miniature fans, listening to tiny emerald birds, and standing by impossibly small but tinkling fountains.

"Is it not beautiful?" said the Emperor. "If you asked me what I have done here, I could answer you well. I have made birds sing, I have made forests murmur, I have set people to walking in this woodland, enjoying the leaves and shadows and songs. That is what I have done." "

But, oh, Emperor!" pleaded the flier, on his knees, the tears pouring down his face. "I have done a similar thing! I have found beauty. I have flown on the morning wind. I have looked down on all the sleeping houses and gardens. I have smelled the sea and even seen it, beyond the hills, from my high place. And I have soared like a bird; oh, I cannot say how beautiful it is up there, in the sky, with the wind about me, the wind blowing me here like a feather, there like a fan, the way the sky smells in the morning! And how free one feels! That is beautiful, Emperor, that is beautiful too!"

"Yes," said the Emperor sadly, "I know it must be true. For I felt my heart move with you in the air and I wondered: What is it like? How does it feel? How do the distant pools look from so high? And how my houses and servants? Like ants? And how the distant towns not yet awake?"

"Then spare me!"

"But there are times," said the Emperor, more sadly still, "when one must lose a little beauty if one is to keep what little beauty one already has. I do not fear you, yourself, but I fear another man."

"What man?"

"Some other man who, seeing you, will build a thing of bright papers and bamboo like this. But the other man will have an evil face and an evil heart, and the beauty will be gone. It is this man I fear."

"Why? Why?"

"Who is to say that someday just such a man, in just such an apparatus of paper and reed, might not fly in the sky and drop huge stones upon the Great Wall of China?" said the Emperor. No one moved or said a word.

"Off with his head," said the Emperor. The executioner whirled his silver ax.

"Burn the kite and the inventor's body and bury their ashes together," said the Emperor. The servants retreated to obey.

The Emperor turned to his hand-servant, who had seen the man flying. "Hold your tongue. It was all a dream, a most sorrowful and beautiful dream. And that farmer in the distant field who also saw, tell him it would pay him to consider it only a vision. If ever the word passes around, you and the farmer die within the hour."

"You are merciful, Emperor."

"No, not merciful," said the old man. Beyond the garden wall he saw the guards burning the beautiful machine of paper and reeds that smelled of the morning wind. He saw the dark smoke climb into the sky. "No, only very much bewildered and afraid." He saw the guards digging a tiny pit wherein to bury the ashes. "What is the life of one man against those of a million others? I must take solace from that thought."

He took the key from its chain about his neck and once more wound up the beautiful miniature garden. He stood looking out across the land at the Great Wall, the peaceful town, the green fields, the rivers and streams. He sighed. The tiny garden whirled its hidden and delicate machinery and set itself in motion; tiny people walked in forests, tiny faces loped through sun-speckled glades in beautiful shining pelts, and among the tiny trees flew little bits of high song and bright blue and yellow colour, flying, flying, flying in that small sky.

"Oh," said the Emperor, closing his eyes, "look at the birds, look at the birds!"

From *Golden Apples of the Sun* Doubleday, 1953

14. What is the major conflict in the story?
 - A. Between the inventor who wanted to live, and the Emperor who ordered him to be killed
 - B. Between progress and tradition
 - C. Between the emperor and himself whether to allow his servant to live
 - D. Between the servant and the emperor whether to look at the invention, or to ignore it
15. Who does the Emperor believe should be responsible for an invention and why?
 - A. Society is responsible for an invention, because they may use it for a different purpose
 - B. An evil man is responsible for an invention, because he will use it in the wrong way
 - C. The emperor himself is responsible for all inventions, because he needs to protect the citizens of his country
 - D. The inventor should be responsible for his invention, because his talent should be used only for good
16. The Emperor suggests that creative talents should be used to
 - A. build airplanes
 - B. make elaborate toys
 - C. tear down walls
 - D. discipline servants
17. Some people could believe that "The only circumstance in which we are justified in taking the life of another person is in self-defense." Would the Emperor agree with this statement and why?
 - A. The emperor would disagree, because he thinks he can kill whoever he wants
 - B. The emperor would disagree, because he is only concerned with his own power
 - C. The emperor would agree, because he killed the inventor to protect future generations
 - D. The emperor would agree, because he did not kill the executioner
18. The Emperor did not rush out to see the flying machine when his servant first told him about it because the Emperor
 - A. was too frail to run
 - B. had just awakened from a dream

- C. wanted time to think about what it might mean
 - D. was testing the servant's loyalty to him
19. What does the flying machine symbolize in this story?
- A. Change
 - B. A bird
 - C. Invention
 - D. Happiness
20. Why did the Emperor ask the servant if anyone else had seen the inventor flying?
- A. He wanted more time to awaken, but also wanted to be first to see it
 - B. He wanted to know if word about the invention had gotten around
 - C. He wanted to test the servant's loyalty to him
 - D. He wanted to kill the inventor without any witnesses
21. The Emperor seems to view the Great Wall as a
- A. protector of his way of life
 - B. popular tourist attraction
 - C. symbol of the human spirit
 - D. way to prevent people from escaping
22. Why does the Emperor ask the inventor twice, "What have you done?"
- A. He did not hear the inventor the first time
 - B. He wanted to assert power over the inventor and make him answer twice
 - C. He wanted to draw attention to the possible outcomes of the invention
 - D. He wanted to point out to the servant the words of the inventor
23. Which group of words best helps you to understand the message of this story?
- A. Strength, joy, humor
 - B. Foolishness, anger, endurance
 - C. Communication, friendship, honesty
 - D. Fear, frustration, bewilderment
24. What does the emperor's response to the servant's cry of "a miracle" reveal about his outlook?
- A. The emperor believes that miracles do not occur
 - B. The emperor appreciates what he has and not look for new things
 - C. The emperor believes that miracles are a gift or a dream
 - D. The emperor does not trust his servants
25. Why is the emperor called an old man in paragraph 15?
- A. He is set in his ways
 - B. He is afraid and feels weak
 - C. He is powerful and controlling
 - D. He is sad

26. In paragraph 1 it refers to “too happy” and in paragraph 10, the emperor says the invention is “too beautiful.” How does this reflect on the theme of the story?
- A. Being too happy or beautiful is never a good thing for anyone
 - B. People need to be careful about what they wish for
 - C. Inventions are dangerous
 - D. It is worthwhile to have a little less beauty but stay safe

Passages 3 and 4: (Form A non-accommodated, Form B accommodated)

Passage [3]

Theodore Roosevelt

United States President (1901-1909)

Inaugural Address

Saturday, March 4, 1905

The energetic Republican President had taken his first oath of office upon the death of President McKinley, who died of an assassin's gunshot wounds on September 14, 1901. Mr. Roosevelt had been President himself for three years at the election of 1904. The inaugural celebration was the largest and most diverse of any in memory—cowboys, American Indians (including the Apache Chief Geronimo), coal miners, soldiers, and students were some of the groups represented. The oath of office was administered on the East Portico of the Capitol by Chief Justice Melville Fuller.

MY FELLOW CITIZENS, no people on Earth have more cause to be thankful than ours, and this is said reverently, in no spirit of boastfulness in our own strength, but with gratitude to the Giver of Good who has blessed us with the conditions which have enabled us to achieve so large a measure of well-being and of happiness.

To us as a people it has been granted to lay the foundations of our national life in a new continent. We are the heirs of the ages, and yet we have had to pay few of the penalties which in old countries are exacted by the dead hand of a bygone civilization.... Our life has called for the vigor and effort without which the manlier and hardier virtues wither away.

Under such conditions it would be our own fault if we failed; and the success which we have had in the past, the success which we confidently believe the future will bring, should cause in us no feeling of vainglory, but rather a deep and abiding realization of all which life has offered us; a full acknowledgment of the responsibility which is ours; and a fixed determination to show that under a free government a mighty people can thrive best, alike as regards the things of the body and the things of the soul.

Much has been given us, and much will rightfully be expected from us. We have duties to others and duties to ourselves; and we can shirk neither. We have become a great nation, forced by the fact of its greatness into relations with the other nations of the Earth, and we must behave as befits a people with such responsibilities.

Toward all other nations, large and small, our attitude must be one of cordial and sincere friendship. We must show not only in our words, but in our deeds, that we are earnestly desirous of securing their goodwill by acting toward them in a spirit of just and generous recognition of all their rights.

But justice and generosity in a nation, as in an individual, count most when shown not by the weak but by the strong. While ever careful to refrain from wrongdoing others, we must be no

less insistent that we are not wronged ourselves. We wish peace, but we wish the peace of justice, the peace of righteousness. We wish it because we think it is right and not because we are afraid. No weak nation that acts manfully and justly should ever have cause to fear us, and no strong power should ever be able to single us out as a subject for insolent aggression.

Our relations with the other powers of the world are important; but still more important are our relations among ourselves. Such growth in wealth, in population, and in power as this nation has seen during the century and a quarter of its national life is inevitably accompanied by a like growth in the problems which are ever before every nation that rises to greatness. Power invariably means both responsibility and danger. Our forefathers faced certain perils which we have outgrown. We now face other perils, the very existence of which it was impossible that they should foresee.

Modern life is both complex and intense, and the tremendous changes wrought by the extraordinary industrial development of the last half century are felt in every fiber of our social and political being. Never before have men tried so vast and formidable an experiment as that of administering the affairs of a continent under the forms of a democratic republic. The conditions which have told for our marvelous material well-being—which have developed to a very high degree our energy, self-reliance, and individual initiative—have also brought the care and anxiety inseparable from the accumulation of great wealth in industrial centers.

Upon the success of our experiment much depends, not only as regards our own welfare, but as regards the welfare of mankind. If we fail, the cause of free self-government throughout the world will rock to its foundations, and therefore our responsibility is heavy, to ourselves, to the world as it is today, and to the generations yet unborn.

There is no good reason why we should fear the future, but there is every reason why we should face it seriously, neither hiding from ourselves the gravity of the problems before us nor fearing to approach these problems with the unbending, unflinching purpose to solve them aright.

Yet, after all, though the problems are new, though the tasks set before us differ from the tasks set before our fathers who founded and preserved this republic, the spirit in which these tasks must be undertaken and these problems faced, if our duty is to be well done, remains essentially unchanged. We know that self-government is difficult. We know that no people needs such high traits of character as that people which seeks to govern its affairs aright through the freely expressed will of the free men who compose it.

But we have faith that we shall not prove false to memories of the men of the mighty past. They did their work, they left us the splendid heritage we now enjoy. We in our turn have an assured confidence that we shall be able to leave this heritage unwasted and enlarged to our children and our children's children.

To do so we must show, not merely in great crises, but in the everyday affairs of life, the qualities of practical intelligence, of courage, of hardihood, and endurance, and above all the power of devotion to a lofty ideal, which made great the men who founded this republic in the

days of Washington, which made great the men who preserved this republic in the days of Abraham Lincoln.

Text courtesy of Bartleby Library.

1. According to the introduction, what made Roosevelt's inaugural celebration different from those of the past?
 - A. A wide range of people were in attendance.
 - B. Many politicians were invited to participate.
 - C. His inaugural address was longer than that of any other president.
 - D. It was the first celebration open to anyone who wanted to attend.
2. What does paragraph 7, "But justice and generosity..." reveal about Roosevelt's view of how the United States should relate to foreign countries?
 - A. The US should be fair and peaceful to all nations; it is the responsibility of a powerful country
 - B. The US should be just to all nations; but should not allow itself to be bullied
 - C. The US must let all nations know that fairness means that they will go to war if necessary
 - D. The US must show all countries who wish to make war that peace is most valued
3. Roosevelt begins his address by saying, "no people on Earth have more cause to be thankful than ours." According to Roosevelt, what are the specific reasons the American people at that time had to be thankful?
 - A. They are now a free government but they have suffered less than previous generations
 - B. They have shown the world that they are powerful and only wish peace
 - C. They carry a great responsibility to show the world that democracy can be successful
 - D. They have seen industrial success, although that has come with financial pressures
4. In paragraph 8, Roosevelt refers to the perils the nation faced. He means the nation faced
 - A. critics who questioned its practices
 - B. certain dangers in the past
 - C. rapid population growth
 - D. financial difficulties
5. In paragraph 9, Roosevelt describes governing the United States as formidable. He means that governing was
 - A. a conventional undertaking
 - B. an unexpected development
 - C. a lengthy process
 - D. a difficult challenge
6. Roosevelt emphasizes "responsibility" and "duty" throughout his address. According to Roosevelt, why should the nation take responsibility?
 - A. They have outgrown some of the old perils that have faced the nation in its early days
 - B. They have become more anxious as they have accumulated more wealth
 - C. They need to work together with other nations to ensure world peace

- D. They have been blessed with so much, that it is their duty to do something good with it
7. What are two responsibilities or duties that Roosevelt believed were important?
- A. To become a more diverse society, and to respect all people
 - B. To show that free government can be successful, and to be generous to other nations
 - C. To learn from the history of our past, and to celebrate our country's traditions
 - D. To continue to be successful, and to accumulate wealth so that America can be a superpower
8. What is the experiment that Roosevelt describes in paragraph 10?
- A. The implementation of a democratic form of government on a very large scale
 - B. The successful industrialization of a country with a democratic government
 - C. The resolution of social problems that come with great wealth and power
 - D. The balance between American power and the need for peaceful relations abroad
9. Which of the following best describes Roosevelt's ideas about the relationship between progress and problems?
- A. He believes that in the future progress will not lead to problems.
 - B. He believes progress solves most problems once thought unsolvable.
 - C. He believes a nation cannot have progress without also having problems.
 - D. He believes progress can solve only certain types of problems.
10. Throughout his address, Roosevelt uses persuasive strategies to appeal to his audience. Which is an example of a persuasive strategy Roosevelt used?
- A. Roosevelt really loved America
 - B. Roosevelt gave specific examples of things people can do to make America great
 - C. Roosevelt spoke about the country's strength and tied it to his message
 - D. Roosevelt told a personal story to be persuasive
11. Roosevelt most likely refers to Washington and Lincoln at the end of the address in order to
- A. praise the speaking styles of previous presidents
 - B. encourage listeners to study history
 - C. recall accomplishments from the past
 - D. suggest that government was more powerful in the past
12. When Roosevelt says in paragraph 14, "To do so we must show, not merely in great crises, but in the everyday affairs of life, the qualities of practical intelligence, of courage..." he most likely means:
- A. Street smarts and bravery can be more important than book learning
 - B. Our fine qualities must be shown all the time in our regular living
 - C. Great crises can bring out the best in people
 - D. To get to know what people are like, you should look at them in their everyday living
13. What themes and values did Roosevelt stress in his speech?
- A. Gratitude, responsibility
 - B. Future-mindedness, honesty
 - C. Generosity, self-reliance

D. Self-control, diligence

Passage [4]**The Open Window**

by Saki (H. H. Munro) (1870–1916)

"My aunt will be down presently, Mr. Nuttel," said a very self-possessed young lady of fifteen; "in the meantime you must try and put up with me."

Framton Nuttel endeavored to say the correct something which should duly flatter the niece of the moment without unduly discounting the aunt that was to come. Privately he doubted more than ever whether these formal visits on a succession of total strangers would do much towards helping the nerve cure which he was supposed to be undergoing.

"I know how it will be," his sister had said when he was preparing to migrate to this rural retreat; "you will bury yourself down there and not speak to a living soul, and your nerves will be worse than ever from moping. I shall just give you letters of introduction to all the people I know there.

Some of them, as far as I can remember, were quite nice."

Framton wondered whether Mrs. Sappleton, the lady to whom he was presenting one of the letters of introduction, came into the nice division.

"Do you know many of the people round here?" asked the niece, when she judged that they had had sufficient silent communion.

"Hardly a soul," said Framton. "My sister was staying here, at the rectory, you know, some four years ago, and she gave me letters of introduction to some of the people here."

He made the last statement in a tone of distinct regret.

"Then you know practically nothing about my aunt?" pursued the self-possessed young lady.

"Only her name and address," admitted the caller. He was wondering whether Mrs. Sappleton was in the married or widowed state. An undefinable something about the room seemed to suggest masculine habitation.

"Her great tragedy happened just three years ago," said the child; "that would be since your sister's time."

"Her tragedy?" asked Framton; somehow in this restful country spot tragedies seemed out of place.

"You may wonder why we keep that window wide open on an October afternoon," said the niece, indicating a large French window that opened onto a lawn.

"It is quite warm for the time of the year," said Framton; "but has that window got anything to do with the tragedy?"

"Out through that window, three years ago to a day, her husband and her two young brothers went off for their day's shooting. They never came back. In crossing the moor to their favorite snipe-shooting ground they were all three engulfed in a treacherous piece of bog. It had been that dreadful wet summer, you know, and places that were safe in other years gave way suddenly without warning. Their bodies were never recovered. That was

the dreadful part of it." Here the child's voice lost its self-possessed note and became falteringly human. "Poor aunt always thinks that they will come back someday, they and the little brown spaniel that was lost with them, and walk in at that window just as they used to do. That is why the window is kept open every evening till it is quite dusk. Poor dear aunt, she has often told me how they went out, her husband with his white waterproof coat over his arm, and Ronnie, her youngest brother, singing, 'Bertie, why do you bound?' as he always did to tease her, because she said it got on her nerves. Do you know, sometimes on still, quiet evenings like this, I almost get a creepy feeling that they will all walk in through that window—"

She broke off with a little shudder. It was a relief to Framton when the aunt bustled into the room with a whirl of apologies for being late in making her appearance.

"I hope Vera has been amusing you?" she said.

"She has been very interesting," said Framton.

"I hope you don't mind the open window," said Mrs. Sappleton briskly; "my husband and brothers will be home directly from shooting, and they always come in this way. They've been out for snipe in the marshes today, so they'll make a fine mess over my poor carpets. So like you menfolk, isn't it?"

She rattled on cheerfully about the shooting and the scarcity of birds, and the prospects for duck in the winter. To Framton it was all purely horrible. He made a desperate but only partially successful effort to turn the talk onto a less ghastly topic; he was conscious that his hostess was giving him only a fragment of her attention, and her eyes were constantly straying past him to the open window and the lawn beyond. It was certainly an unfortunate coincidence that he should have paid his visit on this tragic anniversary.

"The doctors agree in ordering me complete rest, an absence of mental excitement, and avoidance of anything in the nature of violent physical exercise," announced Framton, who labored under the tolerably widespread delusion that total strangers and chance acquaintances are hungry for the least detail of one's ailments and infirmities, their cause and cure. "On the matter of diet they are not so much in agreement," he continued.

"No?" said Mrs. Sappleton, in a voice which only replaced a yawn at the last moment. Then she suddenly brightened into alert attention—but not to what Framton was saying.

"Here they are at last!" she cried. "Just in time for tea, and don't they look as if they were muddy up to the eyes!"

Framton shivered slightly and turned towards the niece with a look intended to convey sympathetic comprehension. The child was staring out through the open window with a dazed horror in her eyes. In a chill shock of nameless fear Framton swung round in his seat and looked in the same direction.

In the deepening twilight three figures were walking across the lawn towards the window; they all carried guns under their arms, and one of them was additionally burdened with a white coat hung over his shoulders. A tired brown spaniel kept close at their heels. Noiselessly they neared the house, and then a hoarse young voice chanted out of the dusk: "I said, Bertie, why do you bound?"

Framton grabbed wildly at his stick and hat; the hall door, the gravel drive, and the front gate were dimly noted stages in his headlong retreat. A cyclist coming along the road had to run into the hedge to avoid imminent collision.

"Here we are, my dear," said the bearer of the white mackintosh, coming in through the window; "fairly muddy, but most of it's dry. Who was that who bolted out as we came up?"

"A most extraordinary man, a Mr. Nuttel," said Mrs. Sappleton; "could only talk about his illnesses, and dashed off without a word of goodbye or apology when you arrived. One would think he had seen a ghost."

"I expect it was the spaniel," said the niece calmly; "he told me he had a horror of dogs. He was once hunted into a cemetery somewhere on the banks of the Ganges by a pack of pariah dogs, and had to spend the night in a newly dug grave with the creatures snarling and grinning and foaming just above him. Enough to make anyone lose their nerve."

Romance at short notice was her specialty.

14. Which of the following best describes what happens in the story?
 - a. A young man visits his aunt and tells her about a recent tragedy.
 - b. A young girl amuses her family by telling them scary stories.
 - c. A young girl makes up a story and frightens a nervous visitor.
 - d. A family plays a trick on a young man from out of town.
15. In paragraph 1, the narrator describes the niece as a "**self-possessed** young lady." This means the niece is
 - a. confident and in control
 - b. very imaginative
 - c. well dressed and polite
 - d. very talkative
16. Why does Framton Nuttel visit the countryside?
 - a. To cure his nervous condition
 - b. To take a tour of the area
 - c. To meet his sister's friends
 - d. To visit his aunt, Mrs. Sappleton
17. Why is Vera interested in whether Framton knows anything about her aunt?
 - a. She is being polite and friendly to a guest
 - b. She wants to gossip about her aunt with him
 - c. She wants to see if she can play her trick on him
 - d. She wants to know if she can trust him
18. According to Vera, why does Mrs. Sappleton leave the window open every day?
 - a. She hopes her husband and brothers will return through it one day
 - b. She likes to hear the singing before the men return each day
 - c. She likes the fresh night air

- d. She keeps it open so that there is another exit out of the room
19. What is Framton's reaction when Mrs. Sappleton begins to talk about her husband and brothers returning from their hunting trip, and why does he react this way?
- a. He shivers with fear because of his nervous condition
 - b. He thinks it is horrible because she thinks they are still alive
 - c. He begins to talk about his illness because he needs her sympathy
 - d. He becomes more friendly with Vera because they are close in age
20. One critic described Saki as an author who uses both comedy and horror in his writing. How does the critic's description apply to "The Open Window"?
- a. The story has comedy when Framton talks about his illness, and has horror in the death of the hunters
 - b. The story shows horror and comedy because Vera is a horrible person to play such a trick
 - c. The story is comic because there are ridiculous people in it, and horrible because people are mean
 - d. The story begins in horror with a story of death, but turns to comedy when we find it is not true
21. In paragraph 8, the narrator says Framton has the **delusion** that total strangers are interested in his nervous condition. This means Framton
- a. has a common illness related to his nerves
 - b. is worried about what people think of him
 - c. dreams that he is constantly being watched
 - d. has a mistaken idea about how others view him
22. How can Vera's character be viewed as both cruel and harmless?
- a. She is cruel to play a prank on a nervous man, but is harmless because she just made up a story
 - b. She is cruel because she drove him out of the house, but is harmless because he was a bad guest
 - c. She is cruel because she enjoys a story about death, but is harmless because she is still young
 - d. She is cruel because she lies about Framton, but she is harmless because no one knew him
23. What is the main strategy the author uses in the story?
- a. He creates conflict among the important characters.
 - b. He adds romance to the relationship between the two main characters.
 - c. He withholds information from the reader until the story's conclusion.
 - d. He creates suspense by disguising the identity of a character.
24. What does Vera do or say to make Framton believe her?
- a. She is very caring for her aunt and acts like her aunt is deluded
 - b. She tries to be amusing and entertaining

- c. She refers to the story being written about in the newspaper
 - d. She uses true details like the coat and the song
25. When it says, “Romance at short notice was her specialty,” which words could replace the word “romance” while keeping the intended meaning of the sentence?
- a. Romantic love
 - b. Elaborate stories
 - c. Nasty gossip
 - d. False information
26. What was Framton’s initial feeling about visiting people he did not know?
- a. He was anxious to meet new people and make new friends
 - b. He hoped that there would be someone around his age
 - c. He worried that it would not help with the cure for his nerves
 - d. He was excited to meet the people who were so nice to his sister

Appendix D

Training Scripts

Training Script for Accommodated Passages:

Please enter your Study ID number:

You will be taking a multiple choice test. This is a practice and training session to teach you how to use the features of the computer.

The passage is on the top of the screen and contains the story or passage to be read. You can scroll up and down the passage using these buttons. (show/highlight these features)

Click **next** (picture of button) to move to the next question.

Click **back** to go back to a question.

Click **mark** to mark a question and come back to it later.

Click **review** to see how many questions you have marked and how many you have not completed.

On the bottom of the screen is the **time remaining** for this section. You will have 25 minutes for each section (screen shot)

If you need more time when time is up, click the “**Extra time**” button. (screen shot)

If you would like to have any part of the test read aloud to you, click on **this button** found at the end of every paragraph.

This is a sample question. Take time to scroll up and down the passage, choose an answer and hear the question read aloud to you.

Sample Passage:

What are Frogs?

Frogs are a kind of small animal belonging to a group called vertebrates (animals with backbones) known as amphibians. This means that they live part of their life in water and the other part of it on land. Amphibians are cold-blooded animals. This means that their body temperature is the same as the surrounding temperature. Their skin absorbs water into their body so they do not have to drink water to survive.

Sample Question:

What is this passage about?

- A. Frogs
- B. Stars
- C. Animals
- D. Furniture

The test will begin when you click the **begin test** button. If you are ready to begin, click the **begin test** button.

Training Script for Non-Accommodated Passages:

Please enter your Study ID number:

You will be taking a multiple choice test. This is a practice and training session to teach you how to use the features of the computer.

The passage is on the top of the screen and contains the story or passage to be read. You can scroll up and down the passage using these buttons. (show/highlight these features)

Click **next** (picture of button) to move to the next question.

Click **back** to go back to a question.

Click **mark** to mark a question and come back to it later.

Click **review** to see how many questions you have marked and how many you have not completed.

On the bottom of the screen is the **time remaining** for this section. You will have 25 minutes for each section (screen shot)

This is a sample question. Take time to scroll up and down the passage, choose an answer and hear the question read aloud to you.

Sample Passage:

What are Frogs?

Frogs are a kind of small animal belonging to a group called vertebrates (animals with backbones) known as amphibians. This means that they live part of their life in water and the other part of it on land. Amphibians are cold-blooded animals. This means that their body temperature is the same as the surrounding temperature. Their skin absorbs water into their body so they do not have to drink water to survive.

Sample Question:

What is this passage about?

- A. Frogs
- B. Stars
- C. Animals
- D. Furniture

The test will begin when you click the **begin test** button. If you are ready to begin, click the **begin test** button.