2020

THE EFFECT OF THE READ-ALOUD AND EXTENDED TIME ACCOMMODATIONS ON NAEP FOURTH AND EIGHTH GRADE READING AND MATHEMATICS FOR STUDENTS WITH DISABILITIES

Ida Tam

THE EFFECT OF THE READ-ALOUD AND EXTENDED TIME ACCOMMODATIONS ON NAEP FOURTH AND EIGHTH GRADE READING AND MATHEMATICS FOR STUDENTS WITH DISABILITIES

A dissertation submitted in partial fulfillment

of the requirements for the degree of

DOCTOR OF EDUCATION

to the faculty of the

DEPARTMENT OF ADMINISTRATIVE AND INSTRUCTIONAL LEADERSHIP

of

THE SCHOOL OF EDUCATION

at

ST. JOHN'S UNIVERSITY

New York

by

Ida Tam

Date Submitted _____ Date Approved _____

_____ _____

Ida Tam Dr. Rene Parmar

ABSTRACT

THE EFFECT OF THE READ-ALOUD AND EXTENDED TIME ACCOMMODATIONS ON NAEP FOURTH AND EIGHTH GRADE READING AND MATHEMATICS FOR STUDENTS WITH DISABILITIES

Ida Tam

The purpose of this study was to examine the effects of the read-aloud and extended time testing accommodations on the fourth and eighth grade National Assessment of Educational Progress (NAEP) in reading and math for students with disabilities. This non-experimental study used the 2013 reading and math NAEP, restricted data set, to examine the fourth and eighth grade students with disabilities with test accommodations compared to the students with disabilities who did not receive testing accommodations. This study had a large sample size and is considered to be representative of the performance of the student population within the nation. The effectiveness of the read-aloud and extended time accommodations have yielded mixed results when applied to other standardized assessments. The present study compared matched samples of students who did and did not receive the accommodation to evaluate its impact. Further, it will extend the present state of knowledge on the topic related to reading assessment into the area of mathematics assessment. Multiple regression

analyses were used to determine the effectiveness of testing accommodations. Result from this study found that students with disabilities benefitted from the read-aloud accommodation. The extended time accommodations appeared to have benefitted the 4th grade students and not the 8th grade students. Gender was another variable that was explored; however, there was no conclusive findings as the statistical software used in this study did not provide an analysis of interaction effects, which is one limitation of this study. Nonetheless, this study provided some valuable information for future researchers interested in understanding the effectiveness of test accommodations.

ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation to my dissertation committee for their time, dedication, feedback, and guidance. A special thank you to my mentor, Dr. Rene Parmar, who devoted countless hours guiding me throughout the various stages of my dissertation. Your structured, yet gentle push throughout the process was what made the completion of the dissertation possible. You have kept me on track and remained focus the entire time. Dr. James Campbell, for your guidance and directions on the use of the NAEP data. Your calm demeanor made things less stressful. Your wealth of ideas certainly guided me with the pursuit of a topic that I am passionate about. Finally, Dr. Erin Fahle, my greatest regret was not having been enrolled in one of your classes; however, you have provided me with valuable feedback throughout the process. You have raised questions that required me to expand my thought process.

To my father and Uncle Henry and his family for supporting me throughout this journey. Despite unforeseen circumstances, you all have taken a load off my plate as I was working toward completion of my dissertation. I also believe that this process was not possible if, my late mother, was not looking down from above and giving me strength all these years. While I miss you, you are always a part of me.

Finally, Dr. Mark DiBuono, you have been there from the beginning, for which I am forever grateful. Your unconditional support has made this journey possible. Robert Laurenzano, you have pushed me toward a goal that I did not know was possible. You gave me the confidence and strength from the beginning. You believed in my abilities and encouraged me to move forward. Lastly, Esther Morell, for the time I have come to know you, you have been a role model, as a supervisor and mentor. Your wisdom, drive

and approach have provided me with the strength to complete this journey. Your military style did not go unnoticed, along with your weekly check-ins during each stage of the process was greatly appreciated. You helped me brainstorm topics and got me started on writing my first sentence and assured me that things will start flowing. You gave me the confidence on days where I was stuck and lifted my spirits.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

For many years, students with disabilities were often excluded from high stakes testing. It was not until the Individuals with Disabilities Education Act Amendments of 1997 (IDEA 97) that states were required to document how students with disabilities were to participate in high stakes testing programs at the state and national levels. Thurlow and Johnson (2000), based on a review of literature, reported that when students with disabilities were excluded from high stakes testing, there are several unintended consequences. For instance, there was the likelihood of an increase in the referral for special education services, as well as increased rates of retention in grades prior to the testing grades. Further, these students may miss out on valuable instruction that was being provided to students without disabilities since they were not expected to take the accountability assessments. Consequently, the assumption was that the academic expectations for the students with disabilities were lowered (Thurlow & Johnson, 2000).

In 2019, approximately 90% of students classified as having disabilities were included in the national assessment, based on a policy established by the National Assessment Governing Board in 2010. The policy was adopted for NAEP, with the provision of various accommodations based on principles of universal design and consistent with prior research on test accommodations for students with disabilities. A Decision Tree was established for use of accommodations based on students' Individualized Education Plans (IEPs), along with mandates for disaggregated reporting of scores. The most used accommodations for students with disabilities were breaks

during testing and extended time, with provision of aides, separate locations, use of braille, and read aloud as other options.

Students with disabilities have historically performed less well on standardized assessments than their general education peers, even in cases where they are on grade level (NCES, 2017). The overall performance data on students with disabilities in the National Assessment of Educational Progress (NAEP) reported a 31-point gap in reading and a 29-point gap in mathematics at fourth grade, and a 39-point gap in reading scores and a 41-point gap in mathematics at eighth grade (NCES, 2017). To date there has been no systematic research on the usefulness or impact of test accommodations regarding changing student outcomes. While the reports indicated an increase in access to the assessments, it was not clear whether students with disabilities were positively impacted with regard to their test performance.

Purpose of the Study

The purpose of the study was to examine if two of the most commonly used accommodations, reading the test questions aloud and extended time would result in better scores on the NAEP reading and mathematics test for students with disabilities compared to those students who did not receive the accommodation. Both reading and math tests from the NAEP 2013 data were used. Students with disabilities comprised approximately 12% of total students tested, and the majority received test accommodations (65% at Grade 4 and 72% at Grade 8), with extended time being the most frequently provided accommodation.

It was not until the No Child Left Behind Act (NCLB) of 2001 that all school districts who applied for Title I funds were required to provide assurance that they

participate in the biennial state level NAEP testing in reading and math in the 4th and 8th grades (Epstein, 2005). NAEP, is also known as the Nation's Report Card, has been collecting data on student performance since 1969. It is the only assessment that provides the most reliable trend about the performance of the nation's students (Epstein, 2005). From its inception, NAEP's goal was to measure academic achievement at the national level, as well as, measure academic trends by looking at demographic subgroups. Although, NAEP does not report individual student data, it tracks educational progress over a period (Hombo, 2003). As part of the NCLB requirements, NAEP testing now includes students with disabilities and students with limited English proficiency (LEP). Therefore, states and districts are expected to report scores for students with disabilities and schools are held accountable for meeting the Adequate Yearly Progress (AYP) of the subgroups on the state's academic standards (Cook et. al., 2010).

This study addressed a social issue regarding students with disabilities. Students with disabilities were often excluded from competitive academic programs based on their performance on high stakes assessments. Supports provided to assist students during the testing situation do not often result in improved student completion of assessments or improved performance. This may be due to mismatches between student learning needs and accommodations that were provided. Accommodations policies are detailed in most high stakes assessments; however, debate continues regarding implementation, appropriateness, and effects of accommodations on assessment results for students with disabilities (Lazarus 2009).

Ketterlin-Geller (2007) reported that state and federal legislation do not provide a framework for making accommodation and modification decisions. Therefore, selection

of accommodations is based on IEP team recommendations. General education and special education teachers are often confused in terms of how to implement the accommodations and identifying what procedures constitute an accommodation compared with a modification. The National Assessment Governing Board (2010) has developed a policy document on test accommodations for students with disabilities and English Learners which provides a decision-tree for accommodations based on students' IEPs that is now in place for NAEP, which relies on local IEP committee recommendations.

Laws and Policies

The Rehabilitation Act of 1973, the American with Disabilities Act (ADA) of 1990, and the Individuals with Disabilities Educational Act (IDEA) of 1991 and 1997 have played a significant role in understanding testing accommodations for students with disabilities. The Rehabilitation Act (Section 504) of 1973 established a requirement for any program that receives financial assistance where individuals with disabilities are afforded with equal access and participation (Pitoniak & Royer, 2001). Consequently, the act prohibits tests that discriminate against individuals with disabilities. The significance of IDEA 1997 in implementing testing accommodations is reflected in ensuring that individuals with disabilities can obtain employment and education, since jobs may require individuals to obtain credentials that are predicated on passing certain exams. Schools require students to be successful in State achievement exams before obtaining a diploma of graduation and moving toward higher level of education.

IDEA specifies that students with disabilities be provided with the same access to public education as those students without disabilities. A student with disability is

entitled to an individual educational program (IEP) that specifies a plan and services to ensure that students with disabilities can be educated in the least restrictive environment. Since assessments are a major component in the educational system, IDEA ensures that students with disabilities are included in the general state and district assessments. Prior to IDEA 1997, students with disabilities were often excluded from large scale assessments. Reasons for exclusion were attributed to (a) parents not wanting their children to be exposed to possible failure, (b) school personnel feeling that it is not appropriate or worthwhile, and (c) administrators not wanting their overall test performance affected. The 1997 Amendments to the IDEA mandates the inclusion of students with disabilities in participation of the large-scale assessments to ensure that they receive the same educational benefits as those students without disabilities. States are now held accountable for the performance of the students with disabilities.

Test accommodations are changes in the testing conditions that do not alter the construct that is being measured. This is distinguished from test modifications, which changes the construct that is being measured (Royer & Randall, 2012; The Standards of AERA, APA & NCME, 1999). Rieck and Dugger Wadsworth (2005) added that accommodations should be individualized to meet the student's needs and are not to be generically applied to all students with disabilities. Therefore, test accommodations have been highly confusing and controversial, especially when state policies vary on the types of test accommodations allowed. Nonetheless, Royer and Randall (2012, p. 140) provided the four common kinds of test accommodations. Within each accommodation, there are subcategories that specify what the accommodation would look like. The accommodations are:

1. Presentation
    a. Oral (reading test aloud)
    b. Paraphrasing
    c. Technological (computer vs. paper-based presentation)
    d. Braille or large print
    e. Sign Language
    f. Cueing
    g. Spelling Assistance (spell checkers)
    h. Use of manipulatives (using counters on math test)
2. Timing
    a. Extended time
    b. Multiple days or sessions (extending test time over days or sessions)
    c. Separate sessions (separate sessions for sections of a single test)
3. Response
    a. Use of scribes
    b. Use of booklet versus sheet
    c. Marking task book to maintain place allowed
    d. Transcription
4. Setting
    a. Separate room

Research on test accommodations has been difficult and unreliable for several reasons. Pitoniak and Royer (2001) stated that much research has been based on small sample sizes due to not having a large enough population of a disability to conduct an

analysis. In addition, there are great variability among the individuals tested, which make it difficult to classify the examinees into groups for research purposes. The accommodations are adjusted to specific needs of the individuals and therefore not applied consistently. Finally, there is great variability in the type of accommodations that are available. Although test accommodations, when used appropriately, will allow students with disabilities to be educated with their peers, the overuse of accommodations essentially can create a more restrictive environment for the student (Lovett and Lewandowski, 2015).

Further, some have expressed ethical concerns with test accommodations. Lovett and Lewandowski (2015) identified five ethical concerns that are often raised regarding test accommodations. The first concern is related to accommodations being used to benefit students who are already doing well. Lichtenberg (2004) considered testing accommodations, obtained through dubious disability diagnoses, to be one of the ways in which "the academically rich get richer" (p. 19). Essentially, the affluent students have families who are familiar an aware of the services available and are willing to try to utilize every resource they can to give their children an advantage. A second concern is whether the accommodations can "mask deficiencies in assessment programs" (Lovett & Lewandowski, 2015, p. 35). This suggests that if students with disabilities are exempt from taking a test, then one will not have to consider whether the test is appropriate to measure the skills for those students who are not disabled. In this situation, one way to remedy the concern is by asking why the accommodations are necessary for the students with disability and whether the testing procedure is also not appropriate for the nondisabled students as well. A third concern is related to accommodations that can

discriminate against nondisabled students, who would benefit from the accommodation; however, are unable to get the accommodations. A fourth concern is related to the accommodations providing a false representation of competence by an increase in score performance, when, the student lacks the competence in the subject being tested. This concern typically pertains more to certification and licensure exams as opposed to K-12 exams. The last concern discussed is related to the limitations that accommodations can create for students, where students become dependent on certain accommodations, such as extended time, that when these students leave the educational system, they do not have the skill set needed to complete a task under standard conditions. In certain professions, it is often unrealistic to allow an individual extended time in completing the require job responsibilities (Lovett & Lewandowski, 2015).

Ultimately, one needs to recognize that an accommodation is only "an adjustment, to an activity or setting", which removes the barrier for the person with a disability to ensure that this person can "have access equal to that of a person without a disability" (Byrnes, 2000, p.22). Further, the accommodation does not "guarantee success or a specific level of performance" (Byrnes, 2000, p. 22). Rather, the accommodation allows for equal access and a chance for equal participation.

Theoretical Framework

The Individuals with Disabilities Education Act (IDEA, 1975) defined a specific learning disability as "a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations." This disability category includes such conditions as

perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia (a type of language disorder). Based on the federal definition, and subsequent research on diagnoses of learning disabilities, difficulties with information processing remains one of the key definitional components of learning disability. Difficulties with decoding text and reading comprehension are some of the most prevalent forms of difficulty manifest by individuals with learning disabilities. Similarly, a student with attention deficit hyperactivity disorder (ADHD) also has cognitive deficits in processing material. Although, both disabilities classifications are different; however, they do have similar characteristics in that their cognitive functioning are negatively impacted due to their disability.

According to the U.S. Department of Education policy, based on the Reading First Initiative (2004), that is part of the No Child Left Behind Act, (2001), reading is defined as follows:

The term reading means a complex system of deriving meaning from print that requires all the following:

(A) The skills and knowledge to understand how phonemes, or speech sounds, are connected to print.

(B) The ability to decode unfamiliar words.

(C) The ability to read fluently.

(D) Sufficient background information and vocabulary to foster reading comprehension.

(E) The development of appropriate active strategies to construct meaning from print.

(F) The development and maintenance of a motivation to read.

Based on the above definition, Johnstone et al. (2008) conducted a study that analyzed the state reading standards in terms of understanding what students are required to do in the reading assessments. Four modes of interaction identified were visual, tactile, auditory, and multi modal as modes of interacting with printed material (Johnstone et al., 2008). The visual mode of print interaction is known as reading. Individuals will look at print and decipher meaning from symbols. Accommodations can be using enlarged print or through use of computer monitors. Consequently, the ability to interact with the printed material is through reading with the eyes (Johnstone et al., 2008). The tactile print interaction is the use of Braille. The auditory mode of print interaction includes various ways to convert the printed material into audible words and messages, such as books on tape, talking books, text-to-speech programs, or the use of human readers. Finally, the Multi-Model of interaction with print involves a combination of visual, tactile, auditory modes of interaction. For the purposes of this study, the interest is in the Auditory mode of print interaction. Johnstone et al. (2008) analysis found five themes that emerged from their analysis of the state standards that are expected of the students to master.

Theme 1: Reading defined as the acquisition of specific skills (phonemic knowledge, fluency, sight word recognition, and literal comprehension).

Theme 2: Reading defined as the knowledge of conventions or elements in print (rules, language patterns, and structures of the text).

Theme 3: Reading defined as an interactive, thinking activity (interpreting and responding to the material that is in print).

Theme 4: Reading defined as a problem-solving tool (seeking out, organizing, and using information from a variety of sources).

Theme 5: Reading defined as a catalyst for personal growth (vocabulary development, socialization into literary communities, individualized purposes, and reflecting upon self and culture) (p. 224).

Johnstone et al. (2008) found that with the exception of Theme 1, where a student who is receiving auditory accommodations would not be able to demonstrate that he/she can read print fluently as in the phonetic standards, the student with a learning disability is able to demonstrate knowledge and comprehension of texts for the other four themes through the use of auditory presentation. Consequently, Johnstone et al. (2008) study would suggest that the read aloud test accommodation would be an effective accommodation for students with learning disabilities.

Byrnes (2000) stated that students with learning disabilities have a difficult time decoding print. However, they can comprehend the material when they hear it and are able to reason by listening to the material. Therefore, if the reading assessment is to determine if a student can read, one knows they have difficulty with reading. However, when the assessment is to determine if student has mastered content knowledge, the read aloud accommodation would be something that removes the barrier for students with learning disabilities.

For students with ADHD, their deficiency lies in a slower processing speed. In addition, some of the symptoms identified by the American Psychiatric Association for ADHD are distractibility and difficulty sustaining attention. Barkley (1997) stated that students with ADHD often have deficits in their executive functioning skills, seen as a

lack a sense of time and ability to use time management properly. Further, these students typically struggle with initiation of tasks. Consequently, both students with a learning disability and/or ADHD have problems in their cognitive functioning that would interfere with their ability to demonstrate the level of knowledge and competence within the structure of timed tests and may benefit from extended time for completion (Lewandowski et al., 2007).

Even though certain disabilities would appear to warrant some form of accommodation for the student with a disability when it comes to taking standardized exams, the challenge that exists is how there is no uniform consensus across the states in how and what type of accommodations should be provided. Some states allow more standard accommodations, whereas, other states may restrict the accommodations that are being provided. For example, New York State Department of Education (2018) provides guidelines, based on existing disability identification and classification research, for testing accommodations, including a framework for when to use what type of accommodation. Figure 1 indicates the specific conditions for which reading a test aloud is considered to be an acceptable accommodation:

Figure 1.

Reading Aloud Testing Accommodations for Students With Disabilities NYSED (2018). *

Student Characteristics	**Possible Effect on Test-taking**	**Possible**
Processes written information at a slow rate	May not be able to complete exam within standard timeframe.	• Extended time

Student Characteristics	Possible Effect on Test-taking	Possible
	May become fatigued and/or distracted.	• Separate setting • ***Directions read*** • ***Tests read aloud*** • Text-to-speech software (with reading rate adjusted to accommodate processing delays) • Multiple day
Difficulty following/ understanding directions	May not understand what the test requires them to do.	• ***Directions read orally*** • Directions simplified • Additional examples of directions provided
	May have difficulty remembering directions.	• ***Directions reread for each page of***
Difficulty with reading	Reading skills below grade level of test.	• ***Oral reading of tests*** • Text-to-speech software
	Slow reading pace.	• Extended time • Present test reading
Difficulty with auditory processing	Difficulty remembering and/or understanding oral directions.	• ***Repeat directions more than standard number of times*** • Directions simplified • Provide written directions • Preferential seating • Repeat listening
Difficulty with math processing/computati	Unable to memorize basic math facts.	• Use of calculator • Chart of basic math

Note: Adapted from NYSED (2018). Categories without reading aloud accommodations have been removed; presence of reading aloud is highlighted.

It is noted that there are variations in the language used for recommendations, which vary from reading directions to reading actual items. There are also no specifics given as to how frequently a student may have directions read, nor wait time between reading and expected student response. Nevertheless, the framework presented in Figure 2 guides the analysis of the data from the NAEP assessments, where modifications follow

the guidelines presented in the National Assessment Governing Board Policy Statement (2010)

Figure 2

NAEP Nation's Report Card Accommodations Table

Standard NAEP Practice, for SD and/or ELL

Standard Accommodations for SD/ELL Students	***Math***	***Reading***	***Science***	***Writing (CBA)***	***Civics, Economics, Geography, U.S. History***	***Music and Visual Arts***	***TEL (pilot 2013)***	***Writing (before CBA; paper and pencil)***
Extended time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Small group, or one-on-one	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
One-on-one	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Directions only read aloud in English	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Test items read aloud in English – occasional or most/all	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Breaks during test	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Writes directly in the booklet	No	No	No	No	No	No	No	No

Note: NAEP Guidelines for Inclusiveness taken from https://nces.ed.gov/nationsreportcard/about/accom_table.aspx

It is also important to note that as of this time, NAEP does not allow the reading test to be read aloud, except for the questions. This is despite research by Samuels (2014) who reported that a study of 2,000 students in the 4th and 8th grade in California, who took the NAEP 2013, showed a significant difference in their reading scores, when they

had the entire test read aloud to them. Reportedly, the scores were on average, 5.8 points higher than those who did not receive the read aloud accommodation for the 4th graders. However, the difference was smaller for the 8th grade students, but nonetheless, their scores were still higher compared to those students who did not receive the read aloud accommodation. Some have argued against the read aloud accommodation on the reading test because it becomes a "listening test" (Samuels, 2014, p. 17).

Conceptual Framework

Bolt and Thurlow (2007) stated that one of the challenges in the implementation of statewide assessments and accountability measures lies in how to effectively include students with disabilities. Accommodations provided to students with visual and hearing impairment are much more clear-cut and not as controversial as providing accommodations for students with other disabilities, such as learning disability or ADHD. Most would not dispute that for a student with a visual impairment, accommodations may include braille format or large print to allow the student to demonstrate achievement in knowledge of the content material. However, when one looks at a student with a learning disability or ADHD, student accommodations needs are not so evident, and are further complicated with the individual variation that exist for these students. While federal legislation requires that "appropriate" accommodations be provided to students with disabilities for state and district-wide assessment, the legislation does not provide guidance in how to determine whether a specific accommodation is appropriate (Bolt & Thurlow, 2007, p. 15). There is inconclusive empirical support for some accommodations, such as the read-aloud or extended time accommodations, yet students are being provided with the accommodation. The read-

aloud accommodation is one of the most allowed accommodation on state math exams (Bolt & Thurlow, 2007). The math test typically encompasses word problems that may be difficult for students with specific learning disabilities to read independently; therefore, allowing the test to be read aloud, either by a teacher or an audio cassette will allow the student to demonstrate his/her knowledge of math more effectively. Further, Tindal and Ketterlin-Geller (2004) added that the support for allowing the read aloud accommodations on the math test is based on the premise that 1) most multiple choice questions on the math test require students to have sufficient reading skills and 2) the correlation between multiple choice reading and math tests are .70. Extended time accommodations can be provided as a stand-alone accommodation; however, in practice often students with the read-aloud accommodation are also given extended time accommodations. Therefore, the ability to discern which accommodation is affecting the results of a test is hard to distinguish. Very few studies have looked at any accommodation as a stand-alone accommodation. Cohen et al. (2005) did look at extended time as a stand-alone accommodation and found that the accommodation had little effect on student performance.

Another area of concern regarding test accommodations is related to the threats to construct validity. Pomplun and Omar (2000) cited Messick (1994), who identified two major threats to construct validity as construct underrepresentation and construct-irrelevant variance. Construct underrepresentation refers to when the test content is "too narrow and fails to include important dimensions of the construct" (Pomplun & Omar, 2000, p. 21). Construct-irrelevant variance refers to when the test content is "too broad and contains an excess of reliable variance associated with other constructs" (Pomplun &

Omar, 2000, p. 21). Therefore, when test accommodations are used, the scores can be impacted by construct-irrelevant variance. For a test to be valid, it needs to capture the skills the test measures, as well as, how important each skill is in the total score. Consequently, threats to the validity of test scores can lead to skills having different weights than what was initially intended, or the skill was not intended to be measured by the test (Pomplun & Omar, 2000).

Researchers have used different methods to understand the validity of the read aloud accommodation on nonreading tests that is also applicable to the extended time accommodations. The three commonly identified methods are 1) Differential Boost 2) Measurement Comparability and 3) Item-Level Effects of the Read-Aloud Accommodation (this is only applicable to the read-aloud accommodation).

The Differential Boost method of understanding the validity of an accommodation suggest that a valid accommodation should improve the scores of students with disabilities and not for students without disabilities. Further, others have suggested that there should be a more positive impact on the scores of students with disabilities than for students without disabilities (Fuchs et al., 2000). Consequently, an accommodation would not be valid if both students with and without disabilities benefitted from the accommodation, then it would be deemed unfair to provide an accommodation to only students with disabilities.

Measurement Comparability focuses on whether "an accommodation successfully removes construct-irrelevant variance associated with a disability, it follows that certain test measurement characteristics should be similar for test administrations among accommodated students with disabilities and non-accommodated students without

disabilities" (Bolt & Thurlow, 2007, p. 17). Further, Bolt and Ysseldyke (2008) stated that differential item functioning (DIF) can be used "to examine whether item-level performance is consistent across various student groups" (p. 126). When there is evidence to show a high DIF, then the test may not be measuring comparably across different groups and is potentially biased toward one group. When the DIF is limited, then there is evidence to suggest that the test is measuring "similarly across groups" (Bolt & Ysseldyke, 2008, p. 127).

Item Level Effects include examining if the read aloud accommodation is to eliminate reading difficulties for students when taking the math test to demonstrate true math competence, then the benefit should be seen in problem solving questions on the math test and not on the computation questions (Bolt & Thurlow, 2007).

Significance/Importance of the Study

With high stakes testing being a basis for determining eligibility for federal funding for public schools and student graduation, it is important for policy makers, stake holders, educators and parents understand how to most "effectively and accurately measure the performance of all students while preserving the validity of the test scores" (Lai & Berkeley, 2012, p. 160). Further Fewell (2000) added that the goals of assessment are that "we want to gain valid, reliable, and useful information about children without penalizing them through the limits of our measurement system" (p. 41). Within the field of special education, research on ways to improve test performance will result in improved instruction and services for children with disabilities. Therefore, this current research looked at whether the read-aloud and extended time accommodations are effective accommodations.

The purpose of this study was to look at whether the read-aloud and extended time are effective at boosting the performance of students with disabilities in the NAEP reading and math tests. Prior research studies that have looked at read-aloud accommodations have limitations in their sample sizes. Further, other studies have looked at multiple accommodations within a study. Consequently, the results may or may not be conclusive as to which accommodation was having the effect on the student's performance. Therefore, this study focused specifically on two types of accommodations and its effectiveness by analyzing existing data with a relatively large sample size that is representative of students with special needs across the country.

While the primary focus of this study was on test accommodations, this researcher included a brief preliminary analysis to consider the possibility of an interaction in gender differences in testing accommodations. It appears that at the time of this study, previous research on test accommodations did not look at gender differences and their possible relation to test accommodations. There are considerable research that studied gender differences in academic performance in reading and mathematics, in both general education and special education settings, but little to none specifically regarding test accommodations.

Educational research that looked at gender gaps typically looked at the differences in performance and/or achievement levels between male and female students that primarily focused on females possessing stronger verbal skills and males possessing stronger quantitative skills. There is research to suggest that gender gaps favoring male students in math as they get older are more evident amongst the high performing groups

(McGraw et al., 2006). Further, it appears to be related to test item format with males performing better than females on multiple-choice tests (Reardon et al., 2019).

Rather than looking at the existence of gender differences, there is another group of researchers who believe in the gender similarities hypothesis (GSH). Hyde (2005) conducted meta-analyses to summarize studies that looked at gender differences, where she calculated the effect sizes to determine the magnitude of the gender differences in the studies. Hyde (2005) found that aside from a few areas such as motor behaviors (which encompasses physical strength), aspects of sexuality and aggression are areas that appear to show larger gender differences; whereas, areas related to academics, communication, complex problem solving showed no gender differences or a very small gender difference. Further, Hyde (2005) also found that gender differences fluctuated throughout the life span, where age can play an impact, which contradicted those who argued that gender differences are large and stable.

Like many areas within educational research, gender differences are another area where no agreed upon approach or strategy has been established to ensure that educational outcomes are equalized (McGraw et al., 2006). Some researchers have argued that focus on equalizing the gender differences should be done in the earlier grades. In contrast, other researchers have argued that the research gap on gender differences are often very small and whether this type of research actually does more damage than good by promoting stereotypical behaviors. Some have questioned whether male students are at a more disadvantage since they are more often identified as students who need special education services in comparison to female students. Then, there are those researchers who believe that gender differences cannot be study alone as an

independent variable without consideration of race/ethnicity and socioeconomic status (McGraw et al., 2006). Nonetheless, based on test score analyses, gender gaps do exist, and further research is needed in education to ensure that all students have an equitable chance of academic success. Whether gender is of significance when considering effects of testing accommodations is to be determined; however, it is a variable that can provide some invaluable insight to understanding the different aspects within a student and his/her academic performance.

Connection with Social Justice and/or Vincentian Mission in Education

A provision of the Individuals with Disabilities Education Act (1997) identifies inclusion as educating a student with special needs in the least restrictive environment, if appropriate. This would require a teacher to differentiate the instruction to meet the needs of each student in the classroom. The use of standardized testing is a major component in our education system. For students with special needs, effective testing accommodations will ensure that students with disabilities can succeed as their non-disabled counterparts. According to Stone et al. (2016), "public education is one of the few arenas in American society where the hope of equity and the prospects of societal advancement may be realized by all of our children" (p. 12). Therefore, inclusive education is considered as necessary to ensure social justice. Consequently, test accommodations become a critical area to explore and understand to "level the playing field" (Fuchs et al., 2000 p 2). Test accommodations also offer the opportunity for students with disabilities the option of pursuing higher education to become competitive in the job market.

Stone et al (2016) reviewed John Dewey's view on what Dewey considered as Democracy in Education by providing an overview of the current state of special

education, the understanding of inclusion, and the philosophy in special education. Schools are a social institution, where, it is reflective of what society looks like for students. When students with disabilities are part of an inclusive classroom, they can interact with their non-disabled peers. This interaction fosters an understanding of the differences and diverse needs that exist, which will create a more democratic society.

Conclusion

The present study used the above theoretical and conceptual frameworks to interpret student outcome data from the NAEP assessments by investigating whether reading tests aloud and extended time accommodations respond to the barriers faced by students with learning disabilities and ADHD when engaging in high-stakes standardized assessments. It will seek to extend the understanding of the application of read-aloud and extended time, two of the most widely used accommodations, and extend the research into an examination of the impact of read-aloud in mathematics as well as reading assessments.

Research Questions

Based on the research and theoretical perspectives summarized above, the following research questions were constructed to guide the present study.

RQ1. Do special education students who receive the read-aloud accommodation perform better than students who do not receive the read-aloud accommodation in the NAEP reading or math exams for students in the 4^{th} or 8^{th} grade?

RQ2: Is there a gender difference between students who receive the read-aloud accommodation in the performance on the NAEP reading or math exam in the 4^{th} or 8^{th} grade for students with disabilities?

RQ3: Do special education students who receive the extended time accommodation perform better than students who did not receive the extended time accommodation in the NAEP reading or math exams for students in the 4th or 8th grade?

RQ4: Is there a gender difference between students who receive the extended time accommodation in the performance on the NAEP reading or math exam in the 4th or 8th grade for students with disabilities?

Definition of Terms

Accommodations: Changes to how a test is presented, administered or variations in how test taker responds (NAEP Glossary of Terms). Thurlow et al. (2006), defined it as "tools and procedures that provide equitable instructional and assessment access for students with disabilities" (p. 657-658). Further, accommodations do not change the construct that is being measured (Lazarus et al., 2009).

Adequate Yearly Progress (AYP): The measurement of student progress to the level of proficiency in the state standards in each academic area (Bowen & Rude, 2006).

Item Response Theory (IRT): The relationships between the item responses provided by students and the underlying score scales (NCES, 2019)

Modifications: Changes to the test that involves the change of the construct of the test (Bolt & Thurlow, 2004).

National Assessment of Educational Progress (NAEP): a survey of student achievement that measures academic progress which assesses what students know and can do in various subject areas (reading, math, science, writing, history, geography and the arts) (Hombo, 2003).

Plausible Values: "Proficiency estimates for an individual NAEP respondent, drawn at random from a conditional distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses" (NCES,2019, p.9) These are not individual test scores. They are intermediary computations for calculating summary statistics for groups of students.

Universal Design: Developing educational tests that "entail more flexibility in construction and administration than traditional educational tests" (Royer & Randall, 2012, p. 152). This would suggest that the construction and administration of the tests are more flexible to meet the needs of the different students so it will eliminate the need for test accommodations.

CHAPTER 2

REVIEW OF RELATED LITERATURE

This chapter provides a summary of some of the key research with regard to test accommodations for students with disabilities. The two widely-used accommodations addressed in the present study, read-aloud and extended time, are discussed. The chapter concludes with a brief summary of research on gender differences in allocation and responsiveness to test accommodations.

Research on Implementation of Read-Aloud Accommodations

Read-Aloud in Reading Assessment

Meloy, Deville and Frisbie (2002) examined the effect of the read aloud testing accommodation on students with and without a learning disability in reading. Their study consisted of a sample of 260 students (24% with a learning disability in reading and 76% without a disability) from two middle schools in the midwestern school district. Students were randomly assigned to two experimental testing conditions using four tests of the Iowa Tests of Basic skills. Students with additional diagnosis, such as emotionally disturbed were not included in this study. The study used four ITBS tests: Science, Usage and Expression, Math Problem Solving and Data Interpretation, and Reading Comprehension (Meloy et al., 2002). The study revealed several important issues when deciding on read-aloud accommodations. The authors had developed scripts for the read aloud condition to ensure consistency of administration. A two-way ANOVA showed that there was a significant difference between the mean scores of students in the non-learning disability group and the learning disability group in all four tests. Further, all students scored higher on the tests under the read aloud conditions than under standard

conditions, however, the interaction effects were not significant (Meloy et al., 2002). The research reveals that there may be variations within the implementation of the read-aloud accommodation, despite standardized protocols for administration, that could impact outcomes, as well as, the possibility that reading aloud does not meet the Differential Boost criteria, as all students benefitted.

McKevitt and Elliott (2003) study looked at the effects and consequences of the read aloud testing accommodations on a standardized reading test. Seventy-nine eighth grade students from a midwestern suburban junior high school participated in the study. Forty of the students were identified with a disability and receiving special education services. Thirty-nine were general education students. A total of five teachers participated, where four were special education teachers and one general education English teacher. McKevitt and Elliott (2003) used the TerraNova Multiple Assessment Reading test in this study. Teachers were asked to complete the Academic Competence Evaluation Scale (ACES-Teacher Form) and Assessment Accommodations Checklist (AAC) for each participating student. The AAC was used to create accommodations packages for all students in both groups. Teachers were asked to review the TerraNova and determine which accommodation they believed would be helpful regardless of the disability. Students were randomly assigned to two conditions. Condition One was where students completed one form of the test with no accommodations and one form of the test with teacher-recommended accommodations. Students in Condition Two completed one form of the test with no accommodations and the other form with the read-aloud accommodations in addition to those recommended by the teacher. Results suggested that the most frequently chosen accommodations by teachers were extended time, directions

read aloud and verbal encouragement. Results found that students without disabilities scored significantly higher than students with disabilities on all parts of the test in both testing conditions. Therefore, no differential effect of accommodations on the two groups of students was demonstrated. The teacher-recommended accommodations alone benefitted only the group of students with disabilities, though not significantly, whereas the read-aloud accommodation in addition to the teacher recommended accommodation benefitted both groups. Anecdotally, students with and without disabilities, reported that some liked the read-aloud accommodation, stating that it ensured that they were paying closer attention to the material. However, others stated that they were unable to read ahead or go back if they were to read it independently. Yet, other students found it to be "annoying" because it was either too fast or too slow (McKevitt and Elliott, 2003, p. 596). Ultimately, there was no differential effects for students with disabilities on the read aloud accommodation. Therefore, the McKevitt and Elliott (2003) questioned the validity of the accommodation if both groups benefitted from the accommodation.

Huynh and Barton (2006) looked at the effect of oral accommodations on test structure and student performance on the Reading test of the South Carolina High School Exit Examination (HSSE). This test measured six objectives: 1) decoding and word meaning, 2) main idea, 3) details, 4) analysis of literature, 5) reference usage and 6) inference. Students with mild disabilities took the regular form, while students with severe disabilities took the Oral form. The oral form of the exam consisted of three formats: 1) audiotape operated by test administrator, 2) audiotape operated by student and 3) reading aloud by test administrator. This study included 89,319 students who were categorized into 3 groups: 1) Students with Disabilities who received the ORAL form, 2)

Students with Disabilities who received the regular form and 3) Students without a disability who received a regular form. The disability makeup of the students consisted of N=301 for physical disabilities; N=242 for Emotional disabilities, N=2906 for Learning disabilities and N=287 educable mentally. This is a retrospective study that used factor analysis to determine if there were changes to the internal structure of the test. Further, the researchers used ANCOVA to control for background variables. Results of the study found that the ORAL form did not change the test's internal structure, which meant that the construct of the test remained intact. Students with disabilities who took the ORAL form did as well as students with disabilities who took the regular form.

Laitusis (2010) looked at the impact of audio presentation on a standardized reading comprehension test for 4th and 8th grade students, through the use of a repeated measures design. Students with and without reading-based learning disabilities took both a standard administration and a read-aloud administration of the reading comprehension test. Eighty-four public and private New Jersey schools participated in this study. Gender was evenly distributed. A total of 527 with reading learning disability and 654 without a reading learning disability in the 4th grade participated. A total of 376 students with a reading disability and 471 without a learning disability in the 8th grade participated in the study. Materials used included two equated forms of the Gate-McGinitie Reading Tests (GMRT) fourth Edition (Reading comprehension subtest only) and one form of the Woodcock Johnson III Diagnostic Reading Battery (reading fluency subtest (Laitusis, 2010). The 4th grade students also had two additional subtests from the Woodcock Johnson (Letter-Word Identification and Word Attack). Students were presented with both an audio and standard version of the test. A within-subjects design was used where

all students took the test with and without accommodations. Results of this study found that students with learning disabilities benefitted differently from the read-aloud accommodations in the 4th and 8th grade. A differential boost in performance was seen greater in the 4th grade than the 8th grade, which suggested that listening comprehension becomes stronger of reading ability at the older grade. While there was some evidence to suggest that read-aloud did benefit students with disabilities; however, limitations in this study were also noted. The two populations had different ability distributions, where the students with disabilities were skewed toward the lower end of the scale. In addition, the reading comprehension assessment may not be generalizable to other reading assessments because these passages were shorter than typical reading assessments. The possibility that students' assignment to testing condition, whether it was audio or standard first may have had an effect (Laitusis, 2010).

Read-Aloud in Mathematics Assessment

Huynh et al. (2004) examined the administration of oral accommodations on student performance on the mathematics portion of the South Carolina High School Exit Examination (HSEE). Three groups of students were compared: (1) Students with Disabilities who received the ORAL form, (2) Students with Disabilities who received the regular form and (3) Students without a disability who received a regular form. The sample size consisted of 934 students with disabilities who took the ORAL form, 911 for students with disabilities who took the regular form, and 29,137 students without disabilities who took the regular form. Similar to their study on the oral accommodations for the reading portion of the South Carolina High School Exit Examination (HSSE), this is a retrospective study that used factor analysis to determine whether there were any

changes to the internal structure of the test. An ANCOVA was used to control for background variables. Results of this study found that students who took the oral administration did better than students with disabilities who took the regular form.

In Elbaum's (2007) study, the performance of students with and without a learning disability on a math test with a read-aloud accommodation were compared. Six hundred twenty-five middle and high school students (n=388 with learning disability) were administered a two equivalent 30-item multiple choice tests. The test items used were taken from practice materials for the math statewide tests based on fifth grade standards. A pilot test of form equivalence was administered under a standard condition to a sample of students who were demographically comparable to the participants in this study. Findings showed that the read aloud accommodation improved performance for both students with and without disabilities. Further, students without disabilities showed greater benefit with the accommodations on average than those students with disabilities. Students without disabilities showed a gain of twice the magnitude compared to those students with a disability. In addition, the results of the study showed that the benefits of the oral accommodations differ for students at the elementary level as compared to the secondary level students. For the elementary students, a greater boost was seen in the scores for students with disabilities. In contrast, for the secondary students, the benefit was greater for students without disabilities (Elbaum, 2007).

Helwig et al. (1999) study examined the effect of video accommodation for a standardized mathematics exam. In this study, middle school student's achievement on a traditional administration of a standardized math test was compared with an oral presentation of a comparable version. The authors suggested that computation problems

for students with reading difficulties did not require the necessary reading skills to complete the task; however, in cases of math word problems, students with reading difficulties present a challenge in their abilities to solve the word problems correctly. Terminologies in math are not part of the general vocabulary and are specific to math content, which further contributes to greater problems for students with reading difficulties (Helwig et al., 1999). A total of two hundred forty-seven students in the sixth grade participated in this study. Each student was assessed on (a) Oral reading fluency defined as number of correct words read per minute; (b) Basic mathematics skill – A 21-question test with 19 computation problems and 2 one-step word problems; and (c) a standardized state Mathematics Achievement test. All students answered 60 items in two testing sessions of 30 questions each. One half of the questions were presented in a standard test booklet and the other half was presented via a video monitor. For the video component, students were given a test booklet and each question was displayed on a monitor in front of the classroom and was read aloud as well. Results found that there were small differences when comparing student performance on the 30-item standard and video versions of the math test. Students who had low reading fluency and above-average performance on the mathematics skills test showed the most improvement when questions were read aloud to them. Therefore, consistent with previous studies, students who cannot decode word problems are at a disadvantage when presented with comprehending and solving the problems without questions being read to them. Further, students who had average decoding abilities did not benefit from having questions read aloud to them. Students who had low math skills also did not benefit from having questions read aloud to them. Helwig et al. (1999) found that despite reading being a

significant barrier to problem solving questions on a math test, students who lacked the proficiency in math did not benefit from having questions read aloud, since they were still unable to solve the problems correctly. Therefore, based on Helwig et al. (1999) study, the read aloud accommodations were unnecessary for the majority of the students.

Read-Aloud Accommodation and Score Validity

The read-aloud accommodation is considered one of the more controversial accommodations allowed on tests, where state accommodation policies vary amongst different states. Read-aloud is often allowed on nonreading tests but prohibited on reading tests. Bolt and Ysseldyke's (2006) analysis looked at data from a large-scale achievement test that was part of a statewide assessment program across 3 consecutive years. The purpose of Bolt and Ysseldyke's (2006) study was to examine the belief that the read aloud accommodation is more appropriate on math tests than on reading tests through an analysis of data from a large-scale achievement test. Students data for the reading/language arts were from the 3rd,7th, and 11th grade levels. The math data were from 4th, 8th, and 10th grade levels. Bolt and Ysseldyke (2006) analyzed data for each grade and content area separately. Three student groups were selected for comparison: a) 5000 randomly selected non-accommodated students without disabilities, b) all non-accommodated students with disabilities, and c) students with disabilities who were coded as receiving the read-aloud accommodation with or without additional scheduling and setting accommodations (p. 335). Results from Bolt and Ysseldyke's (2006) analysis found that there were greater proportion of DIF items identified for those receiving the read aloud accommodations on the reading/language arts test than those on the math test. Therefore, the read aloud accommodation was associated with "greater measurement

incomparability" on the reading/language arts test than on the math test (p. 348). Limitations were reported by Bolt and Ysseldyke (2006) that included whether the accommodations were implemented with integrity. Students receiving the read aloud accommodations were the lowest performing group in the analysis, which would not suggest that the results were positively biased by the examiner. Further, the assumption is that the coding was done correctly since the study involved the use of existing data. The possibility that students who received accommodations should not have received accommodations also existed (Bolt and Ysseldyke, 2006). Nonetheless, the result of Bolt and Ysseldyke (2006) analysis suggested that the read aloud accommodation was less appropriate on a reading test than on a math test.

Another controversy with the read-aloud accommodation, specifically for the reading test lies in whether this accommodation changes the construct of the test. Consequently, if the construct of the test is changed, then the read-aloud accommodation is no longer considered an accommodation and would be considered a modification. Cook et al. (2010) study examined whether the read aloud accommodation used in the 4th grade English language arts assessment changed the construct of the test using a factor analysis. Cook et al. (2010) compared the factor structures for the ELA assessment given to 4th grade students by comparing 4 groups of students:

1) Students without disabilities who took the test under standard conditions
2) Students with learning disabilities who took the test under standard conditions
3) Students with learning disabilities who took the test with accommodations specified by their 504 or IEP plans

4) Students with disabilities who took the assessment with a read-aloud accommodation.

Results from Cook et al. (2010) study found that there is support for states' policies of aggregating scores for AYP purposes for students with and without disabilities and for students with disabilities who received accommodations, including the read aloud accommodation. Further, there is evidence that the test has a degree of validity in this study for those who took the test with and without an accommodation. Ultimately, Cook et al. (2010) suggested that the read aloud can be "appropriately considered an accommodation when used with a reading test" (p. 206).

In another study, Fletcher et al. (2006) looked at the read aloud accommodation performance of 3rd grade students, who were identified as students with dyslexia, where decoding was a major problem and compared them to students who were considered average decoders. Students were administered the Texas Assessment of Knowledge and Skills (TAKs) reading assessment. This reading assessment involved three stories of increasing difficulty. Students read a paragraph that contains the title and some illustrative pictures. Responses are multiple choice and "assesses the literal meaning of the passage, vocabulary, and different aspects of critical reasoning about the material in the paragraph" (Fletcher et al., 2006, p. 140). Students were also administered the Letter Word Identification and Word Attack subtests of the Woodcock Johnson III Tests of Achievement. The Woodcock Johnson subtests measures real words and pseudowords. Therefore, this ensures that students decoding skills are impaired. A total of 91 students with dyslexia and 91 students of average decoders participated in this study. Fletcher et al. (2006) believed that in order to not invalidate the TAKS as an assessment of reading,

the three accommodations used were 1) extending the testing session into two blocks to complete, 2) reading the proper nouns to the students and 3) reading stems and possible responses from the comprehension questions to students after they have read the passage independently (p. 141). Fletcher et al. (2006) provided the rationale for the need to extend the TAKS into two sessions to "counteract the effects of fatigue and loss of interest because of word decoding difficulties" (p. 141). The reading of proper nouns to students would "reduce the frustration and fatigue and preventing loss of motivation without invalidating the test" (Fletcher et al., 2006, p. 141). Since the TAKS exam is a measure of reading comprehension and assesses a student's ability to understand the passages read, the reading of the stem would not invalidate the test, rather it should reduce the challenges for those students with decoding difficulties. Results from Fletcher et al. (2006) study found that students with poor decoding skills who received the accommodations had significantly higher TAKS reading scale scores in comparison to those with poor decoding skills who did not receive the accommodation. There was no significant difference between those students who were average decoders who took the exam with accommodations in comparison to the students who were average decoders who did not receive the accommodation.

While No Child Left Behind Act has played a role in terms of testing all students with disabilities, the controversy and questioning of the type of accommodations being provided are still questionable. Some have questioned the fairness of the accommodations, as well as, whether it modifies the construct validity of the test or provide an unfair advantage to students with disabilities (Fletcher et al., 2009). This degree of controversy becomes more complex when it is applied to students in the upper

grades, as the tests can be a determining factor for promotion into the next grade or meeting requirements of High School graduation. Research on accommodations for the upper grades have been scarce and results are mixed when it comes to the read-aloud accommodations.

Fletcher et al. (2009) looked at accommodations for students in the middle school with a reading disability. Similar to their elementary school study in 2006, Fletcher et al. (2009) administered the TAKS exam to 7th grade students, who were considered as poor readers. Students were compared to students who were average readers. The Letter and Word Attack subtests of the Woodcock Johnson was used. The administration of the accommodations was the same as the elementary school study, where the TAKS was broken down into 2 days of administration, the proper nouns, and the reading stems were read aloud to the students. In addition, the Test of Word Reading Efficiency (TOWRE) was administered to the students. This test measures the reading speed. Two subtests from the Kaufman Brief Intelligence Test -2 – Verbal Reasoning and Matrices were also administered to the students. The Verbal reasoning is a test of receptive vocabulary, where students select pictures based on the words that typically reflect general world knowledge. The Matrices subtest is a nonverbal problem-solving test, where students select a response from several possibilities that would demonstrate a relationship (Fletcher et al., 2009). Group Reading Assessment and Diagnostic Evaluation (GRADE), an untimed test that assesses various aspects of reading ability was also used. Fletcher et. al. (2009) used the listening comprehension section of this test. This test is a measure of understanding vocabulary and nonliteral language, as well as, the ability to make inferences. In its initial analysis, results found that the accommodations were effective,

however not considered valid due to the benefits to students without disabilities. While the results for students who were poor readers and received the accommodation had higher scores compared to the students who were poor readers who did not receive the accommodations, there were no significant interactions found. A possible explanation by Fletcher et al. (2009) suggests that the older students are further behind in their age-level expectations that it may be more difficult to "receive enough differential boost to compensate when students are further below grade level" (p. 461).

Research on Implementation of Extended Time Accommodations

Extended Time in Reading Assessment

Lewandowski et al. (2008) studied the benefits of extended time on performance in a reading comprehension test in high school students. Sixty-four students, half with learning disabilities, were given the Nelson Denny Reading Comprehension subtest. In addition, to ensure that the groups were comparable, students were given the Raven Progressive Matrices test that measures general cognitive abilities. The Reading fluency subtest of the Woodcock Johnson Tests of Achievement, Third Edition, was also used to describe the presumed group differences in reading speed, which has been shown to predict comprehension scores (Lewandowski et al., 2008). The researchers used T-tests on several dependent measures and 2x2 mixed model ANOVA. Results of this study found that the nondisabled group benefitted significantly more. Even though students with disabilities did show some benefit in the use of extended time where they were able to attempt more of the test items correctly, their performance did not rise to the level of their nondisabled group's performance.

Lovett and Leja's (2015) study looked at students with ADHD symptoms and whether there were benefits to the extended time accommodation for college level students. A total of 159 students participated in this study. Since processing speed and reading fluency are typically skills that relate to executive functioning and is often used as a justification for extended time, subjects in this study were given a processing speed and reading fluency test. Processing speed measures came from the Wechsler Adult Intelligence Scale, Third Edition that included the two subtests of Digit-Symbol Coding and Symbol Search. To test Reading Fluency, the subtest from Woodcock-Johnson Tests of Achievement, Third Edition, was used to assess subjects' reading rate. The comprehension subtest came from the Nelson-Denny Reading Test (NDRT). Subjects were also given the ADHD current symptoms scale, the behavior rating inventory of executive functioning (BRIEF-A), as well as, the Self-Evaluation of Performance on Timed Academic Reading (SEPTAR). Students completed the battery of measures in small groups that lasted 45 minutes. Lovett & Leja (2015) used Correlation Matrix to analyze the results, where correlations were done separately by gender. Results from this study found that students with more ADHD symptoms did not benefit more from extended time. Perhaps, students with ADHD symptoms did not use their extended time effectively due to their executive functioning problems. Further, they did not find substantial differences by gender. Similar to Pariseau et al. (2010) study, which found that students with ADHD slower their pace when they were given extended time; therefore, did not benefit in utilizing the extended time effectively.

Extended Time in Math Assessment

Cohen et al. (2005) conducted two studies to identify the effect of extended time and content knowledge of ninth grade students who took a statewide math test with and without accommodations. Each study consisted of 1250 students with learning disabilities who received extended time accommodation and 1250 students without learning disabilities. Differential Item Functioning (DIF) is considered to be an important methodology to understanding students with disabilities performance on high stakes test, by making an effort to "match examinees on total test scores or proficiency before examining subgroup differences in item or test performance," which was the premise of the first study (Cohen et al., 2005, p. 226). The results of this study did not conclude that accommodations had an effect. However, in their second study, an exploratory analysis, the researchers found that when students lack skills in specific areas in math, a group difference was evident. Further, accommodations did not improve student performance and neither did student's reading abilities. The underachievement of math scores lies in the students not mastering the math skill. While, accommodations did help those students who already possessed the mathematical knowledge improve their performance, it did not help those students who were deficient in their skills (Cohen et al., 2005).

In another study, Lewandowski et al. (2007) looked at middle school students with ADHD and students without ADHD on the math test under standard and extended time conditions. Participants included 18 male and 9 female students that comprised the ADHD group. The control group included 15 male and 12 female students. All participants were between 10-13 years of age and no significant differences were found in terms of gender. All the ADHD participants had a professional diagnosis of ADHD

and most of the participants were taking medication to address the symptoms of ADHD. Students were administered a subtest of Form A of the Woodcock Johnson Tests of Achievement, Third Edition to obtain a norm referenced measure of student's performance on timed math test. Processing speed was obtained through administering the Wechsler Intelligence Scale for Children, Fourth Edition, that used the Processing Composite scores where students were assessed with the Digit-Symbol Coding and Symbol Search. A Mathematics Calculation Test (MCT) was developed by the researchers. It consisted of three-digit by three-digit addition problems. It was an 18-page test that had 30 math problems on each page for a total of 540 problems. Performance on the MCT test was scored at two intervals - at the end of 12 minutes and at the end of 18 minutes. The researchers looked at the number of items completed correctly, number of items attempted, and percentage of attempted items that were correct (Lewandowski et al., 2007). Findings from this study showed that on the math fluency, the control group outperformed the ADHD group. A group difference was also found on the BRIEF with the ADHD group obtaining significantly higher scores/higher levels of impairment than the control group.

The researchers then used ANOVAs to test the differential boost hypothesis, which stated that the ADHD group should show greater gains from standard to extended time than the control group. Findings suggested that the control group benefitted significantly more from the extended time than the ADHD group; therefore, did not support the differential boost hypothesis. The researcher then looked at whether there was a gain in items above the number of correct items completed in 12 minutes. Findings showed that students with ADHD did not benefit more or less from the extended time

than students without ADHD. When the researchers looked at items attempted, the control group outperformed the ADHD group in both time intervals. When looking at the percentage correct, again, the control group outperformed the ADHD group at both time intervals. Ultimately, this study found that the ADHD group did not perform as well as the control group on measures of processing speed, fluency, and executive functioning. Although the researchers in this study acknowledged that some limitations included a small sample of participants, most students in the ADHD group were taking medication and the inability to control for influences of IQ or achievement; but, the findings provided an understanding of how students with ADHD performed on speeded tests, with and without accommodations (Lewandowski et al., 2007).

Extended Time and Score Validity

Critics of test accommodations often discussed that the use of accommodations provide an unfair advantage to those students who utilize the accommodations. However, studies such as Cohen et al. (2005) shows the contrary that accommodations serve as a "tool to equalize one's access to or demonstration of knowledge" (Cohen et. al., 2005, p. 231). Like the read aloud accommodations, extended time accommodations are also controversial. According to Lovett (2010), critics on extended time accommodations argue that these accommodations are typically obtained by high ability affluent students, which results in an unfair advantage. Some have argued that the extended time accommodation is often provided without considering how the effects of the extra time will impact the test scores interpretation. The validity of a test when extended time is given needs to be considered and explored.

Lovett (2010) identifies four major test types and depending on the purpose of the test, one needs to consider how extended time affects the test interpretation. One type of test identified by Lovett (2010) is the postsecondary admissions test, such as the Scholastic Aptitude Test (SAT) or Graduate Record Examination, GRE, where the scores are interpreted to show the individual's general academic skill and is used to predict future educational performance. These tests yield norm-referenced score interpretations, the score would only have meaning if it is compared to other examinees scores. Another type of tests is the certification and licensure testing. Scores from these tests show the examinees skills and knowledge, and competence in a specific area and how it can be applied to performance in real world settings. The third type of test is the K-12 statewide achievement tests. These tests are used to look at student's academic skills, the success of academic program of the school, or how effective are certain instructional practices. Finally, the fourth type of test is the educational measurement, that are designed by teachers in the classroom. These exams look at students' acquisition of knowledge and skills taught in the class. While students are typically given extended time on all four of these types of tests; however, the validity are not uniform across the tests (Lovett, 2010). Lovett (2010) then describes two features that are important in determining the effects of extended time on validity of a test. The first is the type of score interpretations, whether the test will generate norm-referenced scores, or criterion referenced scores. For norm-referenced scores, scores are used to compare one examinee's performance to another examinee's performance. For criterion-referenced scores, these scores are not dependent on scores of other examinees and typically demonstrate mastery of an area. Therefore, for norm-referenced scores, there should always be comparability between the scores. If

accommodations are provided, then there needs to be evidence of comparability for it to be considered a valid measure.

Munger and Loyd (1991) study looked at the effect of time limits on test performance of handicapped and nonhandicapped fifth grade students. The researchers used the Language Usage and Expression and Mathematics Concepts, Level II from the ITBS in their study which was administered to 222 - 5th grade students, including 6 students who were identified as physically handicapped, 94 students who were identified as learning disabled and 112 had no handicapping condition. All the students in the study had an average range of intellectual functioning. Parallel forms of both the Language Usage and Expression and the Mathematics Concepts were administered to both group of students. One form was timed, and the parallel form was untimed. Results from Munger and Loyd (1991) found that the timing condition did not have an effect for both groups of students. A minimal increase in overall score was seen in both groups of students in the untimed condition. Further, both groups attempted at least 90% of the test on the timed condition, which suggested that the handicapped group was not at a disadvantage. Ultimately, Munger and Loyd (1991) concluded that modification in timing did not have an impact on the student's achievement; therefore, changes to timing conditions in this case was not warranted.

Research on Gender Differences and Test Accommodations

Since research on gender differences and test accommodations are extremely limited and virtually non-existent, this section of the literature review looked at a few of the studies on gender differences that have utilized a large sample set in their analysis, specifically ones that have used the NAEP assessment.

Reilly et al. (2019) study looked at 3 decades of student achievement in the United States in reading and writing from the NAEP assessments to determine if there was a gender difference. Reilly et al. (2019) looked at the performance reading data in the NAEP assessment from 1998-2015, where they had a reading sample size of 3.035 million students in the 4th, 8th, and 12th grade. Like the current study, Reilly et al. (2019) used published data and did not recruit participants in their study. Results from their study found that girls had significantly higher reading scores than boys in every grade. When they looked at grade level and gender comparison, they found that a difference between boys and girls were seen in the upper grades in comparison to the lower grades. Reilly et al. (2019) also looked at gender ratios among students who were considered poor readers and gifted readers. The poor readers, were identified as students who fell below the basic proficiency level and the gifted readers, were identified as students who surpassed the advanced proficiency level on the NAEP assessment. Their findings suggested that there were more boys than girls who were poor readers, where the ratio was 1.54 times as many boys as girls falling below the proficiency level by the 12th grade. In contrast, for the advanced readers, there were almost twice as many girls than boys at the advanced proficiency level by the 12th grade (Reilly et al, 2019).

McGraw et al. (2006) looked at gender differences in NAEP math achievement. Their 3 questions of interest were related to: 1) whether there were changes in the gender gaps amongst US students' math achievement in the NAEP exam between 1990 and 2003, 2) whether gender gaps in achievement vary by mathematical strand, that included race/ethnicity and socioeconomic status, and 3) whether there was a difference in how male and female students compared in their attitudes toward math according to the NAEP

student surveys. These researchers used a t-test to compare achievement means and percentages. Findings from this study showed that although a consistent gap was seen between male and female students on their average scale scores, with males consistently scored higher on average than female students; however, the effect size of these gaps were very small. Further, McGraw et al. (2006) found that the gaps that appeared to be the greatest across the grades were seen at the upper end of the percentile range group of male students. Therefore, a larger number of male students were found to have scored in the advanced and proficient levels compared to the female students. It is also important to note that from their study based on the student surveys, female students did not view themselves as "good at math" or "liking math" to the extent that male students did (McGraw et al., 2006, p. 147).

Regarding gender differences and its relation to test format, Reardon et al. (2018) reported findings from state assessments across the U.S. as well as national assessments, including NAEP. Their research indicated that boys perform better on multiple-choice format items in mathematics as compared to constructed response items, and outperformed females in assessments that are largely multiple-choice in format. The findings were consistent across English Language Arts as well as Mathematics assessments. Conversely, girls performed somewhat better in constructed response items, but not significantly so, particularly in mathematics. The results indicated that large-scale testing programs that are predominantly multiple-choice tend to result in higher scores for males (Reardon et al., 2018).

Conclusion

The above literature review indicates that the impact or benefit of test accommodations remains unclear when applied to large-scale assessments. However, it is evident from national trends in NAEP scores that they do not close the achievement gap between students with disabilities and their general education peer groups. While policymakers and educators have struggled to identify and implement appropriate accommodations, more research is necessary to shed light on aspects of decision-making and application that impact educational outcomes for the target group. The present study adds to the literature on the effects of test accommodations by examining outcomes from students who were provided with the read-aloud or extended time accommodation when completing the NAEP Reading and Math assessments.

CHAPTER 3

METHODS AND PROCEDURES

Research Hypotheses

The hypotheses corresponding to the Research Questions are as follows:

H_{01}: There is no statistically significant difference between students with disabilities who received the read-aloud accommodation and students who did not receive the accommodation in the 4th grade or 8th grade NAEP reading or math assessment in 2013.

H_{02}: There is no statistically significant difference between students with disabilities by gender who received the read-aloud accommodations and students who did not receive the read-aloud accommodations in the 4th or 8th grade NAEP reading or math assessment in 2013.

H_{03}: There is no statistically significant difference between students with disabilities who received the extended time accommodation and students who did not receive the accommodation in the 4th or 8th grade NAEP reading or math assessment in 2013.

H_{04}: There is no statistically significant difference between students with disabilities by gender who received the extended time accommodations and students who did not receive the extended time accommodations in the 4th or 8th grade NAEP reading or math assessment in 2013.

Research Design and Data Analysis

This is a non-experimental study, which used existing data from NAEP to conduct various quantitative analyses to determine whether the read-aloud test accommodation and extended time accommodation have an impact on student performance on the reading and math NAEP exams for 4th and 8th grade students with disabilities. Further, there is a brief analysis on whether gender differences exist amongst students who received the read-aloud or extended time accommodations.

The dependent variables are the 4th grade and 8th grade scores on the NAEP reading and mathematics examinations.

The independent variables are:

accommodation for questions (read-aloud vs. not read-aloud)

accommodation for time (extended vs. not extended)

gender (male and female)

Reliability and Validity/Trustworthiness of the Research Design

The present study utilized data from the NAEP assessment of 2013, and therefore has a large sample size derived from students across the country, meeting the power assumptions of the proposed analyses.

Despite the use of a large data set for this study's analysis, the design itself is ex post facto, as the test accommodations were already determined prior to analysis. While quantitative analyses were utilized, a cause and effect relationship cannot be established due to the correlational nature of the analysis. Further, there was no random assignment of the students to treatment conditions. However, the correlational analysis will shed light on whether or not their received accommodations have an impact on student

performance, and whether they are a reasonable course of action in the case of students with mild disabilities when being assessed on their reading and mathematics performance.

The Sample and Population

The goal of NAEP is to measure academic achievement on a national level and identify trends in progress. NAEP does not evaluate individual student performance or schools. Rather, data are collected according to a "multi-level, multi-stage sampling framework" (Hombo, 2003, p.60). In addition to the item responses, NAEP also collects background data from students, teachers, and schools as part of the analysis.

The sampling procedure for NAEP is representative of students in all schools throughout the state. Schools are first selected and then students are sampled from within the schools regardless of the disability status. Once students are selected, then the appropriate accommodations for students with disabilities may be applied or are excluded. Prior to 1996, NAEP did not allow accommodations for special populations, such as for students with disabilities or English language learners. Guidelines were subsequently created to include special populations. All students with a classification of a disability, or a student with an Individualized Education Program (IEP) can now be included in the NAEP assessment, with 3 exceptions that are permitted to exclude students:

1) The IEP team determined that student could not participate in the NAEP assessment
2) The student's cognitive functioning was so severely impaired that he or she could not participate

3) The student's IEP required an accommodation that NAEP did not permit and the student could not demonstrate knowledge of the subject without the accommodation.

(National Center for Educational Statistics, 2018)

Table 1 presents the number of students included in the data analysis within each subgroup. Overall, the pattern of accommodations given to students were similar for each subgroup. There were more students in the 4th grade sample than 8th grade, and more students receiving extended time as compared to having the test read-aloud.

4th Grade NAEP reading Sample Size

A total of 196,003 students participated in the 4th grade NAEP reading assessment in 2013, with 100,610 male students and 95,393 female students. The number of students classified as having a disability was 27,015, or 13.8% of students tested.

A total of 7,913 male students and 4,259 female students received the read-aloud accommodation, which represented 45.9% of the students identified as having a disability.

A total of 10,259 male students and 5,279 female students received the extended time accommodation, which represented 58.6% of the students identified as having a disability.

4th Grade NAEP math Sample Size

A total of 189,639 students participated in the 4th Grade NAEP math assessment in 2013, with 96,702 males and 92,937 female students. The number of students classified as having a disability was 26,803, or 14.1% of the students tested.

A total of 9,567 male students and 5,184 female students received the read-aloud accommodation, or 57.6 % of the students identified as having a disability. A total of 9,850 male students and 5,132 female students received the extended time accommodation, which represented 55.9% of the students identified as having a disability.

8th Grade NAEP reading Sample Size

A total of 176,270 students participated in the 8th grade NAEP reading assessment in 2013, with 89,991 male students and 86,279 female students. The number of students classified as having a disability was 23,143, or 13.1% of the students tested.

A total of 6,243 male students and 3,455 female students received the read-aloud accommodation, which represented 41.9% of students having a disability. A total of 9,502 male students and 5,068 female students received the extended time accommodation, which represented 63% of students having a disability.

8th Grade NAEP math Sample Size

A total of 172,983 students took the 8th grade NAEP math assessment, with 88,505 male students and 84,478 female students. A total of 22,802 students were classified as a student with a disability, or 13.2 % of students tested.

A total of 7,449 male students and 4,155 female students received the read-aloud accommodation, which represented 50.9% of students having a disability. A total of 9,342 male students and 4,923 female students received the extended time accommodations, which represented 62.6% of students having a disability.

Table 1:

Number of Participants With Disabilities From the NAEP Sample by Grade and Accommodation Type

		SWD w/Accommodations		SWD w/o Accommodations	
		M	F	M	F
4th Grade Reading	Read-Aloud	7,913	4,259	9,605	4,741
	Ext Time	10,259	5,279	7,259	3,721
4th Grade Math	Read-Aloud	9,567	5,184	7,298	3,559
	Ext Time	9,850	5,132	7,015	3,611
8th Grade Reading	Read-Aloud	6,243	3,455	8,533	4,422
	Ext Time	9,502	5,068	5,274	2,809
8th Grade Math	Read-Aloud	7,449	4,155	7,027	3,436
	Ext Time	9,342	4,923	5,134	2,668

Population:

The NAEP results for both the 4th and 8th grade reading and math test are considered to be representative of the performance of the student population within the U.S. for the upper elementary grades as the test is taken by students nationwide from a representative sample. While there may be some individual variations, in general the results of the present research may be generalized to the target population.

Instruments: The NAEP

Assessment data used for this study are derived from the NAEP data set for the 4th and 8th grade reading and mathematics tests from 2013. In terms of establishing content validity, NAEP assessments are developed to represent content-area framework, which are developed by the National Assessment Governing Board (NAGB) (Hombo, 2003). Input is obtained from various stakeholders, such as teachers, subject matter experts, interested individuals and the public. Each subject area that is assessed has a test development committee that selects items to be used on the NAEP assessment. The

assessments are reflective of current educational practices and are reviewed and revised approximately every 10-12 years (Hombo, 2003). Once items are developed, questions go through an extensive review process prior to being used as part of the construct validation process. Items are reviewed by multiple individuals to ensure that they are fair and unbiased. Statistical analysis are utilized to determine if the item is functioning as what it was intended, and differential item functioning (DIF) analyses were completed to remove any unfair items. NAEP uses objective and constructed response items. Performance items range from short answer to extended constructed responses, complex math items, essays, and performance items (Hombo, 2003).

The scoring process for NAEP goes through several channels to support reliability of scoring. Books are scanned, multiple choice items are machine-scored. Images of student responses to constructed-response items are scanned and presented via computer for human raters to score. The professional raters are trained before they can score, and rating reliability is monitored within the assessment, through the use of a second scorer. When scores are below the NAEP reliability standards or a shift in one direction is noted, the item scores are discarded, and the scorers are retrained. The quality control measure in the constructed response ensures that the NAEP scale scores and trends have been designated as trustworthy by independent evaluators (Hombo, 2003).

The NAEP results are based on sampling of students and items. NAEP sample design must meet the following three criteria:

1) Practical
2) Lead to efficient administration procedures

3) Produce acceptable levels of precision for the target statistics. (Hombo, 2003, p. 60).

NAEP uses multistage complex sampling procedures. Student samples are drawn to be representative of the nation in the national samples or state for the state samples. Single grade samples range from approximately 8,000 in the national only subject assessment to 150,000 for combined state and national subject assessments. State samples are approximately 2,500 students (Hombo, 2003).

NAEP assessments are administered to groups of 25-30 students; except for those with accommodations. Multiple subjects are assessed in the same session, as the books are designed with common sets of directions and timing. Since students cannot complete the entire assessment, the total item pool is divided into "blocks" of items (Hombo, 2003, p. 64). The block of items is designed in a way that it is balanced within each block of items and appears equally often in the total sets of books. Typically, there are more than 20 test books per grade per subject (Hombo, 2003).

Gonzalez and Rutkowski (2010) stated that the purpose of large scale assessments are to measure what members of a population "know and can do in a given content domain or whether those member have acquired the skills necessary for performing future life activities" (p.126). The NAEP assessments are considered to be a large-scale assessment. These assessments are unique in that students are administered only a fraction of the available items; yet, the design ensures that there is sufficient content coverage across the population, while "reducing assessment burden for any one student" (Gonzalez & Rutkowski, 2010, p. 126). Gonzalez and Rutkowski (2010) describes in greater detail the use of Matrix sampling of items in large scale assessments. Matrix

sampling “allows us to estimate proficiency distributions of the population, while reducing individual examinee burden and testing time at the school and representing the assessment framework satisfactorily” (Gonzalez & Rutkowski, 2010, p. 126).

After each assessment, NAEP releases and replaces approximately 30% of the items in the pool. The analysis techniques used are based on Item Response Theory (IRT). This means that there is tracking of the multiple book compositions, block positions within books and the varied connections between blocks. Background variables for each student and teacher and school variables must also be matched to the correct students (Hombo, 2003). Furthermore, after the scoring and compiling of the NAEP data, the responses are “weighted according to the sample design and population structure and then adjusted for nonresponse” (Burns et al., 2011, p. 238). According to Burns et al. (2011), the analyses of NAEP data include scaling and estimation. In the scaling phase, “item response theory (IRT) procedures are used to estimate the measurement characteristics of each assessment question” (Burns et al., 2011, p. 238). In the estimation phase, “the results of the scaling are used to produce estimates of student achievement (proficiency)” (Burns et al., 2011, p. 238). Finally, the use of marginal maximum likelihood (MML) methodology is used “to estimate characteristics of the proficiency distributions (Burns et al., 2011, p. 238).

According to Burns et al. (2011), NAEP is able to account for sampling error that can result from the sampling a small number of students and a small number of questions through their use of a complex sampling procedure, known as the jackknife replication procedure. This procedure is used to estimate standard errors. Further, NAEP uses “MML procedures to estimate group distributions of scores” (Burns et al., 2011, p. 242).

NAEP Reading Assessment

The NAEP reading assessment measures reading comprehension skills of students, where students are asked to read passages and answer questions about the passages. The reading framework is based on reading research that measures the student's ability in "understanding written text, developing and interpreting meaning, and using meaning as appropriate to the type of text, purpose and situation" (NAGB, 2012, p.2). Literary texts consist of fiction, literary nonfiction, and poetry. Whereas, informational texts consist of exposition, argumentation and persuasive text, and procedural text and documents. NAEP's reading framework addresses the mental processes of thinking that underlies reading comprehension. It is based on the student's ability to locate and recall, integrate and interpret, and critique and evaluate. Further, the assessment also involves a measure of student's knowledge of vocabulary. Vocabulary questions are considered as integrate/interpret. Prior research has suggested a significant association between students' vocabulary and the ability to comprehend what they read (NAGB, 2012). The purpose of the vocabulary assessment is to determine whether students are able "to understand the meanings of the words the writers use to convey new information or meaning, not to measure readers' ability to learn new terms or words" (NAGB,2012, p. 34).

NAEP Math Assessment

The NAEP math assessment measures students' knowledge and skills to demonstrate the ability to apply their knowledge in problem-solving situations (NAGB, 2012). Like the NAEP reading framework, NAEP has also established a math framework. Each question is designed to measure of the 5 mathematical content areas: 1)

number properties and operations, 2) measurement, 3) geometry, 4) data analysis, statistics, and probability and 5) algebra (NAGB, 2012).

For the math assessment, items are categorized based on complexity (i.e. low complexity, moderate complexity, high complexity). The complexity levels ensure that students are utilizing reasoning, performing procedures, and understanding concepts, or solving problems (NAGB, 2012).

Procedures for Collecting Data

Data were obtained from previously collected data sets of the 2013 fourth and eighth grade NAEP scores in reading and math for students with disabilities. Data were further extracted from the data set to include only students with disabilities who received the read-aloud or extended time accommodations and students with disabilities who did not receive the accommodations. This study utilized data from the NAEP public data set for 2013 from the National Center for Education Statistics by using the NAEP Data Explorer to conduct preliminary analysis that was relevant to this study. Further, in depth analysis used the NAEP restricted data set to conduct regression analyses that was not possible with the public data set.

Accessing Data through the NAEP Data Explorer from the Public Website

This researcher initially accessed the NAEP data through the NAEP Data Explorer public website. The website to access this information is through https://www.nationsreportcard.gov/ndecore/xplore/nde. In the initial access of this website, there is a NCES Data Usage agreement that asks one to either agree to the terms or not agree to the terms. The terms basically describe that the access of the data is for statistical purposes only and that the researcher will not identify the identity of the

subjects or schools. Further the researcher will have to agree to not "link any dataset with individually identifiable data from other NCES or non-NCES datasets" (NCES Data Usage Agreement, 2019).

The public website allows the researcher to create a sign in so one can save the reports that were generated. The NAEP Data Explorer allows a researcher to define a criterion, such as choosing a subject, grade, year, and scales for a particular assessment and then conduct simple statistical tests or create tables and charts. This researcher looked at both the reading and math test for grades 4 and 8 for 2013, along with the Composite scale. The researcher then chose to look at the Major Reporting Groups and selected the following variables: All students, gender, and disability status of students, including those with 504 plans. Finally, the data explorer allows for selection of a maximum of two statistics. This researcher chose average scale scores and standard deviations. Once all these items were selected, a report was created.

Accessing Data through the NAEP Restricted Access Data Set

Prior to the access to use NAEP's restricted data, the researcher submitted an affidavit and obtained approval from the Department of Education/NCES. Once approved, the researcher was able to start accessing the data in a secured office at the university. The researcher did some research on accessing NAEP data from the public website. The researcher then reviewed the NAEP data set of interest which was the 2013 NAEP reading and math data set (44RED and 44MAT, respectively) for both 4th and 8th grade students.

To effectively analyze the NAEP data set, certain procedures and protocols needed to be followed. According to the National Center for Education Statistics, NAEP

data cannot be analyzed by using statistical packages that assume simple random sampling. Reportedly, the use of these statistical packages can result in biased estimates and also result in an underestimation of sampling errors, that can produce incorrect *p* values, which will lead to designation of differences that are statistically significant, when in fact there are no statistical significance (NCES, 2019). Therefore, a special statistical software was used to analyze NAEP data. In this study, the NAEPEX Software and AM Software were used for all analysis.

"NAEPEX software is the data extraction program for choosing variables, extracting data, and generating SPSS, SAS, or Stata control statements necessary for the analysis of NAEP data" (NCES 2019, p.8). NAEPEX is also the Electronic Code Book. AM Software is a statistical software developed by the American Institutes for Research (AIR). This statistical software program is designed to be use with large scale assessments (NCES, 2019). "AM uses SPSS syntax to perform statistical analysis with NAEP data files. This software also generates code that uses plausible values and takes into account complex sampling design in the computation of sampling variance" (NCES, 2019, p.1).

Since the present study looks at the 2013 NAEP data set for the 4th and 8th grade reading and math NAEP assessment, procedures for opening the files are the same for all subsets. Therefore, only access to the 4th grade reading NAEP assessment is described in detail. The data source that was used was the 2013 National reading assessments for grades 4 and 8 and the 2013 National math assessments for grades 4 and 8.

Once the data source is opened, a selection screen appears where variables are selected. Even though for this study, the focus was on the NAEP variables, the school

data were included to correctly match students to schools when using both school-level and student-level variables and to produce a single merged file of variables (NCES, 2019 p . 3). In all NAEP analysis, there are several Preselected Data variables that are automatically included and commonly required for analysis. NAEPEX preselects the following variables in all analysis:

RRPCM1-RRPCM5: Plausible Values, NAEP reading assessment

ORIGWT: Student weight

JKUNIT: Jackknife variance unit

REPGRP1: Jackknife variance stratum

SRWT01-SRWT62: Replicate weights

CENSREG: Census region of the country

DSEX: Gender

FIPS: FIPS state code

The researcher is then able to select variables of interest. For this study, the following variables (Table 2) were selected to look at the read-aloud test accommodation and extended time accommodation:

Table 2

Variables extracted from the NAEP 2013 restricted data set for 4th and 8th grade reading and mathematics.

Field Name	Field Label
DSEX	Gender
IEP	Student classified as having a disability (504)
XS04901	Extended time (allowed for all subjects)
XS04907	Test items read aloud in English-Occasional, most or all the time

Once all the variables were identified and selected, one was able to generate a syntax file. For this study, the syntax file that was generated was a SPSS file. Once the SPSS syntax file was generated, the SPSS file was then imported into the AM software program. From this point forward, all the analyses were conducted using the AM software program. Prior to the actual analysis, the researcher needed to define the Weight, Cluster, and Strata to reflect the NAEP sample design.

The three variables that needed to be edited are below (Table 3), where one would click on the variable and then click "edit Metadata."

Table 3

Variables edited from the NAEP 2013 restricted data set.

CATEGORY -VARIABLE	SELECTION - METADATA
ORIGWT	Weight
REPGRP1	Strata
JKUNIT	Cluster

The next step was to select the replicated weights, which were SRWT01 through SRWT62 for the present study. The Replication method used was JK2 (Jackknife procedure).

The dependent variables were the Plausible NAEP reading value composites 1 through 20 for the reading assessment and the Plausible NAEP math value composites 1 through 20 for the math assessment. The independent variables were test read-aloud, extended time and gender. A separate regression analysis was used for each independent variable to study the data.

CHAPTER 4

RESULTS AND FINDINGS

Descriptive Data Analysis

Prior to the researcher's analysis of the NAEP secure data set, the researcher was able to use the NAEP Data Explorer to conduct some preliminary analysis that was relevant to this study from the National Center for Education Statistics public website, as presented in Table 4.

Table 4
Preliminary NAEP 2013 Descriptive Data for the Participating Sample

		Average Scale Score	SD
4th Grade Reading	National	222	37
	Male	219	38
	Female	225	36
	SWD	184	44
	SWOD	227	33
4th Grade Math	National	242	30
	Male	242	30
	Female	241	29
	SWD	218	32
	SWOD	245	28
8th Grade Reading	National	268	34
	Male	263	34
	Female	273	34
	SWD	232	37
	SWOD	272	32
8th Grade Math	National	285	37
	Male	285	38
	Female	284	35
	SWD	249	37
	SWOD	289	34

**Note:* Table adapted from the following sources: US Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2013 Reading Assessment and Math Assessment.

The results from the preliminary NAEP data access suggested that for the 4th grade reading exam, on average, female students scored above the national average and their male counterparts, with a *M* = 225, compared to the national average scale score *M* = 222 and for male students *M* = 219. On average, students with disabilities had an average scale score, *M* = 184 in comparison to students without disabilities, *M*= 227. For the 4th grade math exam, the national average scale score was *M* = 242. Male students, on average had scored within the national average range, with a *M* = 242. In contrast, female students, on average scored slightly below, with a *M* = 241. Students with disabilities, on average scored below students without disabilities, with a *M* = 218 compared to *M* = 245, respectively.

On the 8th grade reading exam, the national average scale score was *M* = 268. Like the 4th grade exam, on average, the scores for the female students surpassed the national average and the male students. The female students attained a *M* = 273 compared to the national average of a *M* = 268 and for the male students a *M* = 263. Students with disabilities had a *M* = 232 in comparison to students without disabilities who had a *M* = 272.

For the 8th grade math exam, results showed that the national average scaled score was *M* = 285. Similar to the 4th grade exam, male students scored within the national average range, with a *M* = 285 in comparison to the female students who had a scale score of *M* = 284. Students with disabilities had a *M* = 249 and students without disabilities had a *M* = 289.

Ultimately, the results showed that as a group, students with disabilities consistently performed below the national average and had average scale scores below

students without disabilities in both the 4th and 8th grade reading and math exam. It also appeared that female students in both the 4th and 8th grade reading exam had an average scale score that surpassed the national average and their male counterparts. Whereas, the male students' performance on the math test were equivalent to the national average and did slightly better than the female students in both the 4th and 8th grade math exam.

Inferential Data Analysis

The next section addresses the four research questions related to this study. Multiple regression analyses were used to answer the questions. A separate regression analysis was run for each of the independent variables. All the regression analyses were run using the AM Software, a statistical software developed by the American Institutes for Research (AIR). This statistical software program is designed to be use with large scale assessments (NCES, 2019). It reports Jackknife variance that evaluates parameter estimates and standard errors to be significant based on *z* scores. The output table that is generated from AM Software has a different presentation in comparison to the output tables that are generated from SPSS. Therefore, all output tables in this study are reflective of the tables produced from AM Software in that it will report Parameter Estimates and Standard Error in *z* scores, as opposed to SPSS regression analysis that reports *Beta* values.

Hypothesis 1: There is no statistically significant difference between students with disabilities who received the read-aloud accommodation and students who did not receive the accommodation in the fourth grade or eighth grade NAEP reading or math assessment in 2013.

Reading and math results for the 4th and 8th grade will be discussed separately. A regression analysis was conducted for each subgroup and results are summarized in Table 5 below.

Table 5

Results For the Read-Aloud Accommodation

	Parameter Name	Estimate	Standard Error	z Score	p>|z|
4th Gr. Reading	Constant	167.977	0.896	187.406	0
	Read-Aloud	3.219	0.135	23.864	0
4th Gr. Math	Constant	207.381	0.655	316.575	0
	Read-Aloud	2.861	0.13	22.049	0
8th Gr. Reading	Constant	222.971	0.927	240.593	0
	Read-Aloud	1.719	0.155	11.107	0
8th Gr. Math	Constant	238.819	0.857	278.824	0
	Read-Aloud	2.343	0.163	14.368	0

Note: 4th Gr. Reading $R^2 = 0.065$, $F(1,105) = 569.468$; $p = 0$
4th Gr. Math $R^2 = 0.091$, $F(1,83) = 486.151$, $p = 0$
8th Gr. Reading $R^2 = 0.026$, $F(1,116) = 123.356$, $p = 0$
8th Gr. Math $R^2 = 0.049$, $F(1,77) = 206.446$, $p = 0$

4th Grade Reading

Using AM statistics, a regression analysis was estimated by looking at the read-aloud accommodation. Although NAEP does not allow the entire reading test to be read aloud, 6.5% of the variance is explained by the read-aloud accommodation, where $R^2 = 0.065$, $F(1,105) = 569.47$, $p = 0$, which was statistically significant. This read-aloud most likely pertain to the instructions and directions being read-aloud to the students, as opposed to the actual test being read-aloud, since read-aloud is not allowed on the NAEP exam.

4th Grade Math

A regression analysis was estimated by looking at the read-aloud accommodation for students who took the 4th grade math assessment. Findings suggested that 9.1% of the variance can be explained by the read-aloud accommodation, which was statistically significant. For this analysis, the $R^2 = 0.091$, $F\ (1,83) = 486.15$, $p = 0$.

8th Grade Reading

A regression analysis was estimated to look at the read-aloud accommodation for the 8th grade reading assessment. Findings suggested that 2.6% of the variance can be explained by the read-aloud accommodation. The $R^2 = 0.026$, $F\ (1,116) = 123.36$, $p = 0$, which was also considered to be statistically significant. As with the 4th grade assessments, the accommodation was for the questions only.

8th Grade Math

A regression analysis was estimated to look at the read-aloud accommodation for the 8th grade math assessment. Findings suggested that 4.9 % of the variance can be explained by the read-aloud accommodation, which was considered statistically significant with a $R^2 = 0.049$, $F\ (1,77) = 206.45$ $p = 0$.

Read-Aloud Accommodations Conclusion

Overall findings for students with disabilities who received the read-aloud accommodation and those students with disabilities who did not receive the read-aloud accommodation suggested that students with disabilities did benefit from having the read-aloud accommodation. It also suggested that students who took the math exam benefitted more than students who took the reading exam for both grades. Close to 5% of the variance can be explained by the 8th grade math exam in comparison to 2.6% of the

variance in the 8th grade reading exam. Similarly, 9.1% of the variance can be explained in the 4th grade math exam in comparison to 6.5% of the variance in the 4th grade reading exam. It also appeared that the variance for the 4th grade exam was a slightly better predictor than of the 8th grade exam in both reading and math.

Hypothesis 2: There is no statistically significant gender difference between students with disabilities who received the read-aloud accommodations and students who did not receive the read-aloud accommodations in the fourth or eighth grade NAEP reading or math assessment in 2013.

To test this hypothesis, a regression analysis was used, and results are presented in Table 6.

Table 6:

Results For The Read-Aloud Accommodation By Gender

	Parameter Name	Estimate	Standard Error	z Score	p>|z|
4th Gr. Reading	Constant	168.506	1.651	102.077	0
	Gender	-0.389	1.142	-0.34	0.734
	Read-Aloud	3.218	0.134	24.009	0
4th Gr. Math	Constant	216.212	1.257	172.014	0
	Gender	-6.547	0.753	-8.698	0
	Read-Aloud	2.842	0.129	21.975	0
8th Gr. Reading	Constant	215.848	1.661	129.973	0
	Gender	5.25	1.015	5.173	0
	Read-Aloud	1.718	0.155	11.118	0
8th Gr. Math	Constant	247.842	1.608	153.139	0
	Gender	-6.669	0.858	-7.77	0
	Read-Aloud	2.321	0.165	14.088	0

Note: 4th Gr Reading R^2 =0.065, $F(2,104) = 297.997$, $p = 0$
4th Gr. Math $R^2 = 0.100$, $F(2,82) = 265.182$, $p = 0$
8th Gr. Reading $R^2 = 0.030$, $F(2,115) = 70.7619$, $p = 0$
8th Gr. Math $R^2 = 0.057$, $F(2,76) = 149.219$, $p = 0$

4th Grade Reading by Gender

Results from a regression analysis suggested that 6.5% of the variance can be explained by the model that included gender and the read-aloud accommodation, with a $R^2 = 0.065$, $F(2,104) = 297.997$, $p = 0$, which was considered to be statistically significant. Although, gender alone, was not considered to be statistically significant, but appeared that male students scored slightly higher than female students, given that the *z* score value was -0.34.

4th Grade Math by Gender

Results from the regression analysis suggested that 1.0% of the variance can be explained by the model, with a $R^2 = 0.10$, $F(2,82) = 265.18$, $p = 0$, and was considered to be statistically significant. Similar to the 4th grade Reading assessment, male students appeared to have scored slightly higher than the female students, where the *z* score was -8.70.

8th Grade Reading by Gender

Results from the regression analysis suggested that 3.0% of the variance can be explained by the model, with a $R^2 = 0.030$, $F(2,115) = 70.76$, $p = 0$, which was considered to be statistically significant. However, it appeared that both male and female students scored similarly with the read-aloud accommodation.

8th Grade Math by Gender

Results from the regression analysis suggested that 5.7% of the variance can be explained in this model, with a $R^2 = 0.057$, $F(2,76) = 149.22$, $p = 0$, which was also considered statistically significant. Results from this analysis suggested that male students scored slightly higher than the female students, with a *z* score was -7.77.

Read-Aloud by Gender Conclusion

All the regression analyses showed that there was a statistically significant difference existed between students who received the read-aloud accommodations by gender. However, in this study, it was unclear as to whether male or female students had benefitted more from the reading aloud accommodations.

Hypothesis 3: There is no statistically significant difference between students with disabilities who received the extended time accommodation and students who did not receive the accommodation in the fourth or eighth grade NAEP reading or math assessment in 2013.

To test this hypothesis, a regression analysis was estimated, and results are presented in Table 7.

Table 7

Results For The Extended Time Accommodation

	Parameter Name	Estimate	Standard Error	z Score	p>|z|
4th Gr. Reading	Constant	174.958	0.932	187.797	0
	Ext Time	2.525	0.164	15.44	0
4th Gr. Math	Constant	211.301	0.676	312.609	0
	Ext Time	1.97	0.133	14.816	0
8th Gr. Reading	Constant	230.658	0.816	282.741	0
	Ext Time	0.378	0.179	2.113	0
8th Gr. Math	Constant	246.157	0.872	282.195	0
	Ext Time	0.744	0.198	3.754	0

Note: 4th Gr. Reading $R^2 = 0.038$, $F(1,105) = 238.389$, $p = 0$
4th Gr. Math $R^2 = 0.041$, $F(1,83) = 219.501$, $p = 0$
8th Gr. Reading $R^2 = 0.001$, $F(1,116) = 4.46491$, $p = 0.037$
8th Gr. Math $R^2 = 0.004$, $F(91,77) = 14.0915$, $p = 0.0003$

4th Grade Reading

A regression analysis was estimated to look at whether there was a significant difference for students with disabilities who received the extended time accommodation in the 4th or 8th grade NAEP reading assessment. Results showed that there was a statistically significant difference between those students who received the extended time accommodation and those students who did not receive the accommodation in reading for the 4th grade students. The $R^2 = 0.038$, $F(1,105) = 238.39$, $p = 0$, which suggested that 3.8% of the variance in this model can be explained by receiving the test accommodation.

4th Grade Math

A regression analysis was used to look at the 4th grade math exam. Like the 4th grade reading, the results showed a statistically significant difference between the students who received the extended time accommodations and those students who did not receive the extended time accommodation. The $R^2 = 0.041$, $F(1,83) = 219.50$, $p = 0$, which suggested that 4.1% of the variance in this model can be explained by the extended time accommodation.

8th Grade Reading

Results of the regression analysis suggested that there was no statistically significant difference between students who received the extended time accommodation and those students who did not receive the extended time accommodations in the 8th grade reading exam. The $R^2 = 0.001$, $F(1,116) = 4.46$, where $p = 0.04$.

8th Grade Math

Results of the regression analysis on the 8th grade math test was similar to the 8th grade reading test in that there was no statistically significant difference between students

who received the extended time accommodation and those students who did not receive the accommodation. The $R^2 = 0.004$, $F(1,77) = 14.09$, $p = 0.0003$

Extended Time Conclusion

Results on the effect of extended time showed that extended time benefitted the students in the 4th grade reading or math exam and there was no statistically significant difference in the 8th grade reading or math exam. Therefore, extended time did not benefit students in the upper grades.

Hypothesis 4: There is no statistically significant gender difference between students with disabilities who received the extended time accommodations and students who did not receive the extended time accommodations in the fourth or eighth grade NAEP reading or math assessment in 2013.

To test the hypothesis, a regression analysis was estimated, and results are presented in Table 8.

Table 8

Results For The Extended Time Accommodation By Gender

	Parameter Name	Estimate	Standard Error	z Score	p>|z|
4th Gr. Reading	Constant	176.136	1.656	106.364	0
	Gender	-0.874	1.133	-0.772	0.44
	Ext Time	2.524	0.163	15.46	0
4th Gr. Math	Constant	220.67	1.335	165.289	0
	Gender	-7.008	0.741	-9.464	0
	Ext Time	1.972	0.131	15.103	0
8th Gr. Reading	Constant	223.482	1.67	133.807	0
	Gender	5.276	1.044	5.054	0
	Ext Time	0.382	0.178	2.141	0.032
8th Gr. Math	Constant	255.664	1.754	145.795	0
	Gender	-7.109	0.909	-7.82	0

Ext Time	0.748	0.196	3.81	0

Note: 4th Gr. Reading $R^2 = 0.038$, $F(2,104) = 122.108$, $p = 0$
4th Gr. Math $R^2 = 0.052$, $F(2,82) = 167.073$, $p = 0$
8th Gr Reading $R^2 = 0.006$, $F(2,115) = 13.4998$, $p = 0$
8th Gr. Math $R^2 = 0.013$, $F(2,76) = 56.4773$, $p = 0$

4th Grade Reading by Gender

Results for the 4th grade reading exam suggested that when gender and extended time was combined, the results were statistically significant, with $R^2 = 0.038$, $F(2,104) = 122.11$, $p = 0$, which indicated that 3.8% of the variance can be explained. However, gender did not appear to have been a significant contributor to this model, given that $p = 0.44$. Despite that gender was not a significant contributor, it did appear that male students performed slightly higher than female students, as reported in the z Score of -.08.

4th Grade Math by Gender

Results for the 4th grade math exam suggest that it was statistically significant, with $R^2 = 0.052$, $F(2,82) = 167.07$, $p = 0$, or 5.2% of the variance can be explained in this model. Like the reading exam, male students appeared to perform slightly higher than female students, as reported in the *z* score of -9.46.

8th Grade Reading by Gender

Results for the 8th grade reading exam showed that there was no statistically significant difference between gender and students who had the extended time accommodation with an $R^2 = 0.006$, $F(2,115) = 13.50$, $p > .01$.

8th Grade Math by Gender

Results from the 8th grade math exam showed that there was no statistically significant difference between gender and students who had the extended time

accommodation. The $R^2 = 0.013$, $F(2,76)=56.4773$, $p > .05$. However, male students appeared to perform slightly higher than female students, as reported in the z score of -7.82

Gender and Extended Time Conclusion

Results suggested that there was a statistically significant difference in extended time for students by gender in the 4th grade in the reading or math exam and there was no statistically significant difference for students in 8th grade in the reading or math exam. However, like the read-aloud accommodations by gender, one cannot conclude whether male or female students had benefitted more from receiving the extended time accommodations. The reason that one cannot determine which group had benefitted more is because the analysis did not show the interaction effects.

CHAPTER 5

DISCUSSION AND CONCLUSIONS

This chapter will discuss the findings from this study and how they relate to the overall initial questions of whether test accommodations, such as the read-aloud or extended time accommodations are effective predictors in students' performance on the NAEP reading and math exam for students with disabilities. It will look at whether this current study is supportive or refutes prior research in this area. Limitations of this study are included, as well as, recommendations for future practice and research will be discussed.

Implications of Findings

The purpose of this study is to explore whether the read-aloud accommodation and extended time accommodations are effective predictors of performance on the NAEP reading and math exam for students with disabilities, by analyzing an existing data set from the 2013 administration of NAEP. While there is no control in sampling methodologies and individual scores are not reported in the NAEP data, group data is reported and is considered to be representative of the student's across the nation, therefore, the findings of this study can be seen as another attempt in looking at the effects of the two of the most controversial accommodations in the existing literature. This study looks at both reading and math scores for both 4th and 8th grade students who took the NAEP exam in 2013, who were given the read-aloud or extended time accommodation.

Findings for the read-aloud accommodations did not support the Null Hypothesis that there is no statistically significant difference between students with disabilities who

received the read-aloud accommodation and students who did not receive the accommodation in the fourth or eighth grade NAEP reading or math assessment in 2013. The findings from this study suggests that students with disabilities did benefit from receiving the read-aloud accommodation in both the reading and math test. However, the benefit appears to be greater in the math test than in the reading test. Findings also suggest that students in the 4th grade appears to have benefitted from the read-aloud accommodation more than the 8th grade students, based on the variance explained. This finding appears to be reflective with how our academic system is structured. It is more typical to see elementary schools with teachers who engage in read aloud activities in their classrooms more than the upper grades. Marchessault and Larwin (2014) stated that for elementary school students, there is a motivation for these students to read because they are beginning to learn how to read. These students typically enjoy when their teachers read aloud to them. However, for the upper grades, there is the push toward more independent reading, where the interest in reading may decrease for those who are struggling readers once these students realize that they are not comprehending the text like their peers. The upper schoolteachers are often concerned that reading aloud to their students would make their students dependent or passive. However, as reported in Marchessault and Larwin (2014), it is the exact opposite. The important thing is that the teacher can implement various questioning techniques within the read aloud to ensure that the students are actively engage and able to comprehend the material. Read aloud is not a passive activity and is an activity that can continue to help build vocabulary and comprehension amongst the students in the upper grades (Marchessault & Larwin, 2014).

The findings of this study also did not support the second null hypothesis that states that there is no statistically significant gender difference between students with disabilities who received the read-aloud accommodations and students who did not receive the read-aloud accommodations in the fourth or eighth grade NAEP reading and math assessment in 2013. Findings from this study suggest that there is a statistically significant difference between students who received the read-aloud accommodation by gender. In this study, the male students appear to have scored higher than the female students, except for the 8th grade reading exam, which is consistent with national trends for all students (Reardon et al., 2018). However, one is unable to determine if there is a benefit.

The findings of this study shows mixed results regarding the third null hypothesis that states that there is no statistically significant difference between students with disabilities who received the extended time accommodation and students who did not receive the extended time accommodation in the 4th grade or 8th grade reading and math NAEP assessment in 2013. Findings from this study shows that there is a statistically significant difference for the 4th grade students in both the reading and math exam. However, for the 8th grade students, extended time accommodation did not show a statistically significant difference.

Findings for this study also shows mixed results regarding the last null hypothesis that states that there is no statistically significant gender difference between students with disabilities who received the extended time accommodations and students who did not receive the extended time accommodations in the fourth or eighth grade NAEP reading or math assessment in 2013. Findings from this study shows that gender is statistically

significant in the 4th grade exam in both reading and math; however, it is not statistically significant in the 8th grade exam in both reading and math. Again, one is unable to determine whether males or females had more of a benefit with the extended time accommodation.

Relationship to Prior Research

Aron and Loprest (2012) stated that education is "important for all children, but even more so for children with disabilities, whose social and economic opportunities may be limited" discussed how our nation's education system have been servicing students with disabilities (p.97). The evolution of special education had its origins in the civil rights movement in the mid-twentieth century. Federal legislations were established to ensure that children with disabilities have the right to a "free, appropriate public education in the least restrictive setting" (Aron & Loprest, 2012, p. 97). Consequently, the special education system has given students with disabilities greater access to public education. Yet, students with special needs still lag behind their nondisabled peers in the education system (Aron & Loprest, 2012). It is through the No Child Left Behind, where students with disabilities are to be included in state testing by taking the same assessment as their nondisabled peers to determine whether the school is meeting its targeted adequate yearly progress goals toward academic proficiency.

While debate continues to exist on whether students with disabilities should be assessed using the same tests as the nondisabled peers still remain unresolved, where the purpose of this study was not to provide a resolution; however, it does provide some support to previous research that has been conducted on test accommodations. Despite, limitations, which will be discussed later, the findings in this study do suggest that the

continuation of further research in this area is necessary in order to work on closing the achievement gap for students with disabilities. Even though Federal laws are in place, each state is at its own discretion when it comes to recommending test accommodations and assessments. Therefore, the ability to provide clear comparisons and interpretations may be difficult. State assessments typically measure a student's performance based on the state curriculum standards and provide families with individual scores on how a child performs, but it does not allow for comparison with other states. Therefore, it appears that the only standardized test that is administered across the nation is the NAEP assessments. The NAEP assessments allow for certain accommodations to be provided on the reading and math assessments. The present study examines read-aloud and extended time, two widely used and controversial accommodations, for the subgroup of students with disabilities, comparing those who did and did not receive them. At this time, reading aloud on the entire reading assessment is not allowed for NAEP, but only partially for the questions. It is allowed for the math assessment. Extended time is allowed for both reading and mathematics and is the most widely-used strategy.

Relationship to Prior Research Read-Aloud Accommodation

Findings from this study suggest that students with disabilities did benefit from having the read-aloud accommodations in comparison to those students with disabilities who did not have the read-aloud accommodation which is in contrast to the findings of Meloy et al. (2002) or McKevitt and Elliott's (2003) studies. Both prior studies found that there was an overall benefit for all students who received the read aloud accommodation; therefore, there was no differential boost effect seen for students with disabilities. However, one needs to be aware that in both studies, the comparison groups

were between students with disabilities and students without disabilities, which is different from this current studies comparison group that looked at only students with disabilities.

The present study also did not support Huynh and Barton's (2006) study, where they looked at the effect of oral accommodations on test structure and student performance on the Reading test for students who took the South Carolina High School Exit Exam. Huynh and Barton (2006) concluded that students with disabilities who took the test with the oral accommodations did as well as students with disabilities who took the test without accommodations. One may suggest that a reason why this current study did not support Huynh and Barton's (2006) study is that their study consisted of students who were at the end of their High School education career compared to the current studies younger population, who is at the beginning and middle phases of their educational career.

Another significant finding in this current study is that the 4th grade students appears to have benefitted more from the read aloud accommodations than the 8th grade students, which is consistent with Laitusis (2010) and Elbaum's (2007) study. Laitusis (2010) study looked at the impact of audio presentation on standardized reading test for both 4th and 8th grade students for both students with and without a disability. Results from Laitusis (2010) showed that a differential boost in performance was seen greater in the 4th grade than in the 8th grade, which is consistent with the current findings of this study. Similarly, Elbaum's (2007) study looked at the performance of students with and without a learning disability on a math test with the read aloud accommodation for middle school students and high school students. Elbaum (2007) found that there was a

greater boost seen in the lower grades in comparison to the upper grades which was consistent with the findings of this current study.

Bolt and Ysseldyke (2006) looked at data from a large-scale achievement test to examine whether the read aloud accommodation is more appropriate on math tests than on reading tests. Even though, this current study did not look at whether the read-aloud accommodation is more appropriate on the math test over the reading test, this current study does support Bolt and Ysseldyke's (2006) study in that there are greater proportion of DIF items identified for those receiving the read-aloud accommodations on the reading test than those on the math test. Therefore, the read-aloud accommodation appears to be less appropriate for the reading test than the math test. One can make the argument that in this current study, the findings suggests that students benefitted more with the read-aloud accommodation in the math test than in the reading test, with its implication that the read-aloud accommodation for the math test is more appropriate than for the reading test.

Relationship to Prior Research Extended Time Accommodation

Findings from this study found that that there is a statistically significant difference found for students who were given the extended time accommodations for the 4th grade reading and math test only. No significant difference is found for students who received the extended time accommodation in the 8th grade reading and math NAEP exam. The results for the 4th grade exams in this study is not consistent with the findings of any of the previous research on extended time accommodations. One possible explanation is that all the reviewed studies on extended time accommodations looked at the upper grades rather than elementary school students. Perhaps, the benefits seen in the younger grades is that these students just started to take standardized assessments and are

unsure of how to pace themselves, where the extended time would be beneficial. For the upper grades, the skills in the content area may already be lagging to such an extent that the extended time would not be beneficial, as indicted in Cohen et al. (2005), which found that the extended time accommodations had no effect. However, in their exploratory analysis they found that the accommodations did help students who already possessed the necessary skills in the specific content area but did not help those students who did not acquire the needed skills in the first place. Further, Lovett and Leja (2015) and Pariseu et al. (2010) looked at students with ADHD and whether extended time had helped these students. They concluded that there was the possibility that the students with the ADHD symptoms executive functioning skills were already problematic that they were not able to negotiate in the use of extended time effectively. While Lewandowski et al. (2008) study looked at students with disabilities and their nondisabled peers and found that students with disabilities did benefit from the extended time accommodation, their performance did not rise to the level of the performance of their nondisabled peers. However, the students with disabilities were able to complete more test items correctly with the extended time. In this case, then one can make the argument that extended time did benefit students with disabilities in Lewandowski et al. (2008) study; however, their performance was not equivalent to their nondisabled peers would suggest that the skills needed was not necessarily mastered by the students with disabilities.

Relationship to Prior Research When Gender is Considered in Read-Aloud or Extended Time Accommodations

Literature on test accommodations specifically comparing male and female students are scarce and inconclusive currently. Studies that examine whether there are gender differences in students with disabilities who receive either the read-aloud or extended time accommodations are almost non-existent. This researcher conducted a search in various library databases to attempt to locate any previous studies that looks at gender differences and test accommodations. Perhaps this dearth is because there are a higher number of male students placed in special education classes in comparisons to female students. Therefore, there would never be an equitable comparison group.

Gender Differences and Read-Aloud

Findings from the present study shows that male students scored slightly higher when given the read-aloud accommodation on the 4th grade reading and math NAEP exam and the 8th grade math NAEP exam in comparison to the female students. However, for the 8th grade reading NAEP exam, both male and female students had similar scores with the read-aloud accommodation. Some possible explanation for this finding is that there are more male students being tested in comparison to female students. Another possibility, particularly for the math NAEP exam, the male students had consistently benefitted from the read-aloud accommodation on both the 4th and 8th grade exam, which is consistent with McGraw et al. (2006) finding that suggest that as grade level increased the gaps became greater. Therefore, the accommodations did not benefit the female students in this case. Female students tend to have strong verbal skills, where the difference was not seen in the 8th grade reading assessment. The female

students may not have the lack in confidence on the reading test as they probably did on the math test. As stated in McGraw et al. (2006), the female students were found to have a lower level of confidence in their math abilities and interest in comparison to their male counterparts.

Gender Differences and Extended Time

In terms of whether there is a significant difference between male and female students who received the extended time accommodations suggest that for the 4th grade exam, there is a statistically significant difference between the male and female students. Overall, the male students scored higher than the female students in both the reading and math assessment with the extended time accommodation. In contrast, for the 8th grade exam, the findings suggest that there is no statistically significant difference in the reading and math test. However, for the 8th grade math test, the male students did score higher than the female students, despite it being not statistically significant. A possible explanation for these findings again can relate to the number of male students and female students in this sample. Another explanation is that extended time accommodations are often provided to students who have difficulty processing information or have deficits in their executive functioning skills. Typically, students with an ADHD classification are students who would benefit from the extended time accommodation. Research have suggested that there are more male students identified with ADHD than there are female students. Therefore, given this reasoning, the male students would benefit more than the female students. Interestingly, for the 8th grade exam, while there are no statistically significant differences between gender and extended time, the findings did show that male students scored higher than female students on the math exam. Again, this is

consistent with McGraw et al. (2006) study that suggest that the gender gap is greater as the students get older, especially in math.

Limitations of the Study

Some potential threats to the internal validity lies in the possibility that other accommodations may have been recommended in conjunction with the read-aloud or extended time accommodations for some students, as the overlap is not reported in the data set. Often students with disabilities have more than one accommodation. Another threat to the internal validity is the specific disabilities classification of the students. The NAEP data set does not allow for the ability to distinguish between the classifications. One is unable to distinguish whether students with a different classification will benefit or not benefit from the read-aloud or extended time accommodations.

Since this study uses an existing data set, the selection bias and subject characteristics threat is unknown. However, one way that could have improved and reduced the selection bias is with propensity score matching of the subjects. Propensity score matching is a statistical technique for adjusting for selection bias, especially in cases where the researcher has no control over treatment assignments (Lavrakas, 2008). The possible matches that would be created are students IQ scores, student's disability classification, years of receiving special education services, type of special education services currently receiving and socioeconomic status. Once the students are matched and logistic regression can be used, there is the likelihood that there is the reduction in the influence of selection bias in this study. Graham and Kurlaender (2011) indicated that propensity score matching is not without its limitations, but for a large sample size such as NAEP, it is this researcher's belief that propensity score matching can be applied.

Since the NAEP assessment is a large-scale assessment that uses Matrix sampling, "individual measurement precision is sacrificed in the interest of increased content coverage" (Gonzalez & Rutkowski, 2010, p. 126). Therefore, the way the assessment is designed, there is no individual reporting of achievement, rather it is based on group level reporting, extrapolated for each student. This means that one is not able to distinguish whether a specific accommodation is beneficial to individual students.

Although this study found that the read-aloud accommodation is beneficial for the students with disabilities, factors such as the proctors' delivery or presentation of the read aloud could have contributed to the results. Similarly, with the extended time accommodation, the results of this study found that the results are not significant for the upper grades; however, one is not able to identify if there are individual characteristics with this group of students that contributed to the non-significant finding. For example, this school year, may have had a higher number of students with ADHD, where these students did not use the extended time effectively.

Finally, another limitation in this study relates to the use of the AM software in looking at more than one variable and its effect on the accommodations. The AM software is a software program that is designed to be used when analyzing large data sets (NCES, 2019). However, the software program is limited in its capabilities, such that it did not report any interaction analyses.

Recommendations for Future Practice

Based on the findings of this study and what we already know about the laws that are in place for assessment accommodations for students with disabilities, such IDEA or Section 504 of the Rehabilitation Act, as well as, the need for continuous special

education reform, it is imperative that future practice look closely at how to close the achievement gap for students with disabilities by looking at the evidence from national assessments.

Research has shown that teachers have the greatest influence when it comes to creating accommodations for students; however, unfortunately, research also shows that teachers are ineffective in choosing appropriate accommodations that benefit their students (Fuchs et al., 2005). There are a multitude of reasons as to why teachers struggle with choosing the appropriate test accommodations for students, one of which relates to their unfamiliarity of the purpose of each standardize test and what it is supposed to measure. Some studies have found to show that teachers would provide accommodations to students based on demographic characteristics as opposed to the validity of the test. For example, teachers tend to provide accommodations to students of lower socioeconomic status or have lower IQs and reading levels (Fuchs et al., 2005; Fuchs et al., 2000). The present study supports the need to consider a more careful match between student needs and accommodations provided.

To support teachers and the IEP team in making decisions regarding testing accommodations, teachers are encouraged to use the Assessment Accommodation Checklist, developed by Elliott et al. (1998). Thurlow et al. (1995) stated that accommodations guidelines vary amongst the different states in what is allowed and what is prohibited and how scores are treated for students with disabilities, along with the scarce research on the effects of which accommodations are effective on the validity of the standardized tests makes it more complicated for teachers and the IEP team. Elliott et al. (1998) suggested that despite the variability in state guidelines about testing

accommodations, most states will recognize the IEP's recommendation for the accommodations. While the state policies regarding excluding certain accommodations may result in keeping some students with disabilities out of the accountability framework; nonetheless, it is the educator's responsibility to ensure that we are implementing IDEA appropriately and expecting high standards for students with disabilities. Therefore, testing accommodations become crucial for students with disabilities. One way to assist teachers and the IEP team is to use the Assessment Accommodation Checklist (AAC) developed by Elliott et al. (1998) as a starting point. Content validity has been established for this checklist and was field tested with the teachers (Elliott et al., 1998). The AAC is "designed to help organize, record, and encourage teachers to evaluate the helpfulness and fairness of assessment accommodations" that are being provided (Elliott et al., 1998, p. 11). The AAC contains 74 accommodations organized into 8 domains as shown in Figure 3.

Figure 3:

The Assessment Accommodation Checklist: Content Overview and Sample Items

Domain	Descriptions	Example
Motivation	Extrinsic motivators, verbal encouragement	Provide verbal encouragement of effort prior to test
Providing Assistance Prior to Administering Test	Preparation to familiarize students with format	Teach test-taking skills
Scheduling	Flexible scheduling, breaks, some may fatigue more easily, some focus better on certain times of the day	Extended time; Multiple testing days
Setting	Need for minimal distractions, special lighting	Distraction-free (e.g. study carrel); Individual test administration
Assessment Direction	Need modification to directions	Reread directions; paraphrase directions

Providing Assistance During Assessment	Need for scribe; Inability to track test items	Record student responses Turn page for student
Using Aids	Adaptive technology	Pencils adapted in size or grip or electronic reader
Changes in Test Format and Content	Students may b unable to read standard format tests; overwhelmed by number of items on page	Braille or large-type tests, Audiotaped questions

Note: Adapted from "The Assessment Accommodation Checklist: Content Overview and Sample Items" from Elliott, Kratochwill, & Gibertson Schulte (1998).

As stated in Aron and Loprest (2012), those who work in general education and/or special education need to work collaboratively to create solutions to ensure all students are educated without leaving the student's with special needs behind.

Recommendations for Future Research

Future research may want to look more deeply at the read-aloud and extended time accommodations as related to item-level performance for students in the 4th grade and 8th grade NAEP assessments. In this study, it appears that fewer students had an IEP in the 8th grade compared to the 4th grade, and fewer students received the read-aloud and extended time accommodations in the 8th grade. One may suggest that for the upper grades, the assessment measures skills as opposed to the student's ability to decode the material.

Future research may also want to look at making a further distinction between students with disabilities who have an IEP and students who receive accommodations under Section 504. The reason as indicated is that children who receive accommodations through Section 504 typically receives no special education services. Often it is an outside physician who, based on their diagnosis, recommends certain accommodations, where they may not be experts in educational testing accommodations. Therefore, this

leads us to another area of important research related to test accommodations as it relates to who are the people making the recommendations for these accommodations. As indicated earlier, the range of accommodations vary across the states and the characteristics of the students receiving the accommodations also vary.

According to Bielinski (2001), it is the IEP team who has the authority to determine which accommodation the student should receive. However, there is little information regarding how individuals in the IEP team are making their decisions. Bielenski (2001) stated that recommendations for instructional accommodations are not necessarily the same as the recommendations for test accommodations. When one considers test accommodations, there needs to be an understanding of the validity of the test scores to determine whether the accommodation is appropriate (Bielenski, 2001). Further, Fuchs et al. (2005) and other researchers have agreed that an accommodation should be "something that 'levels the playing field' for students with disabilities by allowing them to demonstrate what they are able to do without their competence being obscured by their disabilities" (p.2). A valid accommodation should help student with a disability compensate for his or her disability where the student is able to demonstrate his or her knowledge of the material (Fuchs et al., 2005). There is also consensus that an accommodation should produce a differential boost for the students with disabilities, where the student with disability will show a greater gain in score than a student without a disability if the same accommodation was provided (Fuchs et al., 2005; Phillips, 1994).

Perhaps, the practice of incorporating concepts from the Principles of Uniform Design for Learning is something to consider. In 2010, the U.S. Department of Education funded two large consortia of states to develop assessments that are aligned to the

common core standards (Kettler, 2015). PARCC (2013) and Smarter Balanced (2013) are developing tests to measure student's achievement in English/language arts and math by using assessments through a computer. These computer-based tests are reflective of the principles of Universal Design for Learning, where it recognizes the need for individualized accommodations. Both PARCC and Smarter Balance uses a three-layered system that includes accessibility features for all students (Kettler, 2015). For example, things such as blank paper, flag items for review are incorporated through the computer platform. The ability to change font, text-to-speech software for the math test, the ability to provide responses in a different format are all documented in the student's profile that is based on the IEP (Kettler, 2015). Smarter Balance allows for breaks, passage expansions that are available to all students (Kettler, 2015). Although, this researcher believes that computer adaptive testing can have promising results; at this time one can also see that possible dilemma that it presents, such as schools not having the necessary infrastructure needed to support a building that can have full access to the internet, or a computer for each student. Elliott et al. (2012) stated that "an ideal assessment is one that can be administered under the same conditions across the population, the results of which yield inferences that are equally valid across the range of that population" (p. 129). Therefore, at this time, researchers need to continue their research in test accommodations to ensure that our school's most vulnerable population are included in the school's accountability system and work toward ensuring that educators are working toward closing the achievement gaps. The continued need for evidence-based research is crucial for the success of the students.

One final area that may be of interest to future researchers is through looking at how to improve the ability of the AM software to provide the option of more complex analyses of the data, such as the ability to look at interaction effects of multiple variables. Despite the current limitations of this study, it is the hope of this researcher that this study will lead to future studies that will delve deeper into understanding whether test accommodations have an impact on student's performance.

Ultimately, one needs to keep in mind that accommodations should also not serve to inflate scores of the students with disabilities, as this would be considered unfair and not representative of the outcome (Fuchs et al., 2005). Rather, accommodations are needed to ensure that students with disabilities are given equal access that is like their non-disabled counterparts.

REFERENCES

American Educational Research Association, American Psychological Association, & National council for Educational Measurement. (1999). Standards for educational and psychological testing. Washington, DC: AERA.

Aron, L & Loprest, P. (2012). Disability and the education system. *The Future of Children, 22*(1), 97-122.

Barkley, R. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*(1), 65-94.

Bielinski, J. (2001). Overview of test accommodations. *Assessment for Effective Intervention, 26*(2), 17-20.

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.

Bolt, S.E. & Thurlow, M.L. (2007). Item-level effects of the read-aloud accommodation for students with reading disabilities. *Assessment for Effective Intervention, 33* (1), 15-28.

Bolt, S.E. & Ysseldyke, J. (2008). Accommodating students with disabilities in large-scale testing: A comparison of differential item functioning (DIF) identified across disability types. *Journal of Psychoeducational Assessment, 26*(2), 121-138.

Bolt, S. E. & Ysseldyke, J.E. (2006). Comparing DIF across math and reading/language arts tests for students receiving a read-aloud accommodation. *Applied Measurement in Education, 19*(4), 239-355.

Bowen, S.K. & Rude, H.A. (2006). Assessment and students with disabilities: Issues and challenges with educational reform. *Rural Special Education Quarterly, 25*(3), 24-30.

Burns, S., Wang, X., & Henning, A. (2011). NCES handbook of survey methods. Retrieved from: https://nces.ed.gov/pubs2011/2011609.pdf

Byrnes, M. (2000). Accommodations for students with disabilities: Removing barriers to learning. *NASSP Bulletin, 84*(613), 21-27.

Cohen, A.S., Gregg, N., and Deng, M. (2005). The role of extended time and item content on a high-stakes mathematics test. *Learning Disabilities Research & Practice, 20*(4), 225-233.

Cook, L., Eignor, D., Sawaki, Y., Steinberg, J. & Cline, F. (2010). Using factor analysis to investigate accommodations used by students with disabilities on an English-language arts assessment. *Applied Measurement in Education 23*(2), 187-208.

Elbaum, B. (2007). Effects of an oral testing accommodation on the mathematics performance of secondary students with and without learning disabilities. *Journal of Special Education, 40*, 218-229.

Elliott, S. N., Kratochwill, T. R. & Schulte, A.G. (1998). The assessment accommodation checklist. *Teaching Exceptional Children, 31*(2), 10-14.

Elliott, S.N., Kurz, A., Neergaard, L. (2012). Large-scale assessment for educational accountability. In K.R. Harris, S. Graham, T. Urdan, A.G. Bus, S. Major, & H.L. Swanson (Eds.). *APA Educational Psychology Handbook, Vol.3: Applications to Learning and Teaching*. Washington, DC: American Psychological Association.

Epstein, J. (2005). The national assessment of educational progress: The story of a national test in the making. *American Educational History Journal, 32*(1), 10-19.

Fewell, R.R. (2000). Assessment of young children with special needs: Foundations for tomorrow. *Topics in Early Childhood Special Education 20*(1), 38-47.

Fletcher, J.M, Francis, D.J, Boudousquie, A., Copeland, K., Young, V., Kalinowski, S., & Vaughn, S. (2006). Effects of accommodations on high-stakes testing for students with disabilities. *Exceptional Children, 72*(2), 136-150.

Fletcher, J.M., Francis, D.J, O'Malley, K., Copeland, K., Mehta, P., Caldwell, C.J., Kalinowski, S., Young, V. & Vaughn, S. (2009). Effects of bundled accommodations package on high-stakes testing for middle school students with reading disabilities. *Exceptional Children, 74*(4), 447-463.

Fuchs, L.S., Fuchs, D., & Capizzi, A.M. (2005). Identifying appropriate test accommodations for students with learning disabilities. *Focus on Exceptional Children, 37*(5), 1-8.

Fuchs, L.S., Fuchs, D., Eaton, S.B., Hamlett, C., Binkley,E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children, 67*(1), 67-81.

Gonzalez, E. & Rutkowski, L. (2010). Principles of multiple matrix booklet designs and parameter recovery in large-scale assessments. *IERI Monograph Series: Issues and Methodologies in Large Scale Assessments. www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_03_Chapter_6.pdf*

Graham, S.E. & Kurlaender, M. (2011). Using propensity scores in educational research: General principles and practical applications. *The Journal of Educational Research, 104*, 340-353.

Helwig, R., Rozek-Tedesco, M.A., Tindal, G., Heath, B & Almond, P.J. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *Journal of Educational Research, 93*(2), 113-125.

Hombo, C.M. (2003). NAEP and No Child Left Behind: Technical challenges and practical solutions. *Theory into Practice, 42*(1), 59-65.

Huynh, H. & Barton, K.E. (2006). Performance of students with disabilities under regular and oral administrations of a high stakes reading examination. *Applied Measurement in Education, 19*(1), 21-39.

Huynh, H., Meyer, J.P, & Gallant, D.J. (2004). Comparability of student performance between regular and oral administrations for a high-stakes mathematics test. *Applied Measurement in Education, 17*(1), 39-57.

Hyde, J.S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581-592.

Individuals with Disabilities Education Act of 1991, 20 U.S.C. § 1400 et seq.

Individuals with Disabilities Education Act Amendments of 1997, 20 U.S.C. §1400 *et seq*.

Individuals with Disabilities Education Improvement Act of 2004, Pub.L. 108-446 (2004).

Johnstone, C., Thurlow, M., Thompson, S., & Clapper, A.T. (2008). The potential for multi-modal approaches to reading for students with disabilities as found in state reading standards. *Journal of Disability Policy Studies, 18*(4), 219-229.

Ketterlin-Geller, L.R., Yovanoff, P. & Tindal, G. (2007). Developing a new paradigm for conducting research on accommodations in mathematics testing. *Exceptional Children 73*(3), 331-347.

Kettler, R.J. (2015) Adaptations and access to assessment of common core content. *Review of Research in Education 39,* 295-330.

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

Laitusis, C.C. (2010). Examining the impact of audio presentation on tests of reading comprehension. *Applied Measurement in Education, 23*(2), 153-167.

Lavrakas, P.J. (2008). Encyclopedia of survey research methods. Thousand Oaks, CA: Sage Publications Inc.

Lazarus, S.S., Thurlow, M.L., Laail, K.E., Christensen, L. (2009). A longitudinal analysis of state accommodations policies twelve years of change, 1993-2005. *The Journal of Special Education, 43*(2), 67-80.

Lewandowski, L.J., Lovett.B.J., Parolin, R., Gordon, M., & Codding, R.S. (2007). Extended time accommodations and the mathematics performance of students with and without ADHD. *Journal of Psychoeducational Assessment, 25*(1), 17-28.

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.

Lichtenberg, J. (2004). How the academically rich get richer. *Philosophy and Public Policy Quarterly, 24*(4), 19-27.

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.

Lovett, B.J. & Leja, A.M. (2015). ADHD symptoms and benefit from extended time testing accommodations. *Journal of Attention Disorders, 19*(2), 167-172.

Lovett, B. J. & Lewandowski, L.J. (2015). Testing *accommodations for students with disabilities: Research-based practice*. Washington, DC: APA.

Marchessault, J.K. & Larwin, K.H. (2014). The potential impact of structured read-aloud on middle school reading achievement. *International Journal of Evaluation and Research in Education, 3*(3), 187-196.

McGraw, R., Lubienski, S.T., Strutchens, M.E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. *Journal of Research in Mathematics Education, 37*(2), 129-150.

McKevitt, B.C. & Elliott, S.N. (2003). Effects and perceived consequences of using read aloud and teacher recommended testing accommodations on a reading achievement test. *School Psychology Review, 32*(4), 583-600.

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.

Munger, G.F. & Loyd,B.H. (1991). Effect of speededness on test performance of handicapped and nonhandicapped examinees. *Journal of Educational Research, 85*(1), 53-57.

National Assessment Governing Board (2010). NAEP Testing and Reporting on Students with Disabilities and English Language Learners. Washington DC: National Center for Educational Statistics. https://www.nagb.gov/content/nagb/assets/documents/policies/naep_testandreport_studentswithdisabilities.pdf

National Assessment Governing Board (2012). Mathematics framework for the 2013 National Assessment of Educational Progress https://www.nagb.gov/content/nagb/assets/documents/publications/frameworks/mathematics/2013-mathematics-framework.pdf

National Assessment Governing Board (2012). Reading framework for the 2013 National Assessment of Educational Progress. https://www.nagb.gov/content/nagb/assets/documents/publications/frameworks/reading/2013-reading-framework.pdf

National Center for Education Statistics (2019). Distance learning dataset training. *Data collected through the NAEP.* https://nces.ed.gov/training/datauser/NAEP_03/assets/NAEP_03_Transcript.pdf

National Center for Education Statistics (2019). Distance learning dataset training. *Getting started with NAEP data*. https://nces.ed.gov/training/datauser/NAEP_02/assets/NAEP_02_Transcript.pdf

National Center for Education Statistics (2019). Distance learning dataset training. *NAEP sample design, weights, variance estimation, IRT scaling, and plausible values.* https://nces.ed.gov/training/datauser/NAEP_04/assets/NAEP_04_Transcript.pdf

New York State Department of Education (2018). Testing Accommodations for Students with Disabilities Policy and Tools to Guide Decision-Making and Implementation. Albany, NY: NYSED. http://www.p12.nysed.gov/specialed/publications/documents/testing-accommodations-guide-february-2018.pdf

No Child Left Behind Act of 2001, 20 U.S.C. § *et seq*. (2002)

Pariseau, M.E. Fabiano, G.A., Massetti, G.M, Hart, K.C., & Pelham, Jr., W.E. (2010) Extended time on academic assignments: Does increased time lead to improved performance for children with attention-deficit/hyperactivity disorder? *School Psychology Quarterly, 25*(4), 236-248.

Pitoniak, M.J. & Royer, J.M. (2001). Testing accommodations for examinees with disabilities: A review of psychometric, legal, and social policy issues. *Review of Educational Research, 71*(1), 53-104.

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

Pomplun, M. & Omar, M.H. (2000). Score comparability of a state mathematics assessment across students with and without reading accommodations. *Journal of Applied Psychology, 85*(1) 21-29.

Reardon, S. F., Kalogrides, D., Fahle, E. M., Podolsky, A., & Za´rate, R. C. (2018). The relationship between test item format and gender achievement gaps on math and ELA tests in fourth and eighth grades. *Educational Researcher, 47*, 284–294.

Reardon, S. F., Fahle, E. M., Kalogrides, D., Podolsky, A., Zarate, R. C. (2019). Gender Achievement Gaps in U.S. School Districts *American Educational Research Journal, 56*(6), 2474–2508.

Reilly, D., Neumann,D.I., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the national assessment of educational progress (NAEP). *American Psychologist, 74*(4), 445-458.

Rieck, W. A., Dugger-Wadsworth, D.E. (2005). Assessment accommodations: Helping students with exceptional learning needs. *Intervention in School and Clinic, 41* (2), 105-109.

Royer, J.M. & Randall, J. (2012). Testing accommodations for students with disabilities. *APA Educational Psychology Handbook, 3*, 139-158.

Samuels, C.A. (2014). NAEP provides vehicle for study of read-aloud option: Testing supports are subject of debate. *Education Week, 33*(26), 15-17.

Section 504 of the Rehabilitation Act of 1973, 29 U.S.C. § 701 et seq.

Stone, J.P., Sayman, D.M., Carrero, K. & Lusk, M.E. (2016). Thoughts on Dewey's democracy and (special) education. *Journal of Thought. 50*, 3-17.

Thurlow, M.L. & Johnson, D.R. (2000). High-stakes testing of students with disabilities. *Journal of Teacher Education, 51*(4), 305-314.

Thurlow, M.L. Scott, D.L. & Ysseldyke, J.E. (1995). A compilation of states guidelines for accommodations in assessments for students with disabilities (Synthesis Report # 18). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M.L. Thompson, S.J. & Lazarus, S.S (2006). Considerations for the administration of tests to special needs students: Accommodations, modifications, and more. In S.M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 653-673). Mahwah, NJ: Lawrence Erlbaum.

Thurlow, M.L. Ysseldyke, J.E. & Silverstein, B. (1995). Testing accommodations for students with disabilities. *Remedial and Special Education 16(*5), 260-279.

Tindal, G. & Ketterlin-Geller, L.R. (2004). Research on mathematics test accommodations relevant to NAEP testing. Paper presented at the NAGB Conference on Increasing the Participation of SD and LEP Students in NAEP, Oregon

U.S. Department of Education (2004). Part B student reading skills improvement grants subpart 1 reading first: Sec.1208 definitions. Retrieved from https://www2.ed.gov/policy/elsec/leg/esea02/pg4.html#sec1208

U.S. Department of Education Institute of Education Sciences National Center for Education Statistics (2018, March 13). Digest of Education Statistics. Retrieved from https://nces.ed.gov/programs/digest/d16/tables/dt16_221.12.asp

U. S. Department of Education Institute of Education Sciences National Center for Education Statistics (2018, July 23). Inclusion of Students with Disabilities and English Language Learners. Retrieved from https://nces.ed.gov/nationsreportcard/about/inclusion.aspx

U.S. Department of Education Institute of Education Sciences National Center for Education Statistics (2018, August). The NAEP glossary of terms. Retrieved from http://nces.ed.gov/nationsreportcard/glossary.aspx

U.S. Department of Education Institute of Education Sciences National Center for Education Statistics (2019). NCES Data Usage Agreement. Retrieved from https://www.nationsreportcard.gov/ndecore/xplore/nde

U.S. Department of Education Institute of Education Sciences National Center for Educational Statistics (2019, January 7). Explore the NAEP Reading Results. Retrieved from https://www.nationsreportcard.gov/reading_2017/files/2017_Technical_Appendix_Reading_State.pdf

U.S. Department of Education Institute of Education Sciences National Center for Educational Statistics (2019, January 7). Explore the NAEP Math Results. Retrieved from https://www.nationsreportcard.gov/math_2017/files/2017_Technical_Appendix_Math_State.pdf

U.S. Department of Education Institute of Education Sciences National Center for Educational Statistics (October 2019). Retrieved from https://nces.ed.gov/nationsreportcard/about/accom_table.aspx

Vita

Name	*Ida Tam*
Baccalaureate Degree	*Bachelor of Arts, St. John's University, Staten Island* *Major: Psychology*
Date Graduated	*June, 1996*
Other Degrees and Certificates	*Master of Arts, Teachers College, Columbia University, New York* *Major: Psychological Counseling*
Date Graduated	*June, 1998*
	Master of Education, Teachers College, Columbia University, New York *Major: Psychological Counseling*
Date Graduated	*September, 1998*
	Master of Social Work, New York University, New York *Major: Social Work*
Date Graduated	*June, 2004*
	Advanced Certificate School Building & District Leadership, St. John's University, Staten Island *(2013)*