9-2013

Evaluating the Validity of Accommodations for English Learners through Evidence Based on Response Processes

Katrina Crotts
University of Massachusetts Amherst

EVALUATING THE VALIDITY OF ACCOMMODATIONS FOR ENGLISH LEARNERS THROUGH EVIDENCE BASED ON RESPONSE PROCESSES

A Dissertation Presented

By

KATRINA M. CROTTS

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF EDUCATION

September 2013

Education

Psychometric Methods, Educational Statistics, and Research Methods

EVALUATING THE VALIDITY OF ACCOMMODATIONS FOR ENGLISH LEARNERS THROUGH EVIDENCE BASED ON RESPONSE PROCESSES

A Dissertation Presented

By

KATRINA M. CROTTS

Approved as to style and content by:

____
Stephen G. Sireci, Committee Chairperson

____
Craig S. Wells, Committee Member

____
Jennifer Randall, Committee Member

____
Robert N. Moll, Committee Member

____
Christine B. McCormick,
Dean of the College of Education

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support and guidance of so many people. First and foremost, I would like to thank my advisor and mentor, Stephen Sireci. During my time at UMASS, Steve has given me continuous encouragement and support. He helped me develop as a researcher, critical thinker, and psychometrician. I can't thank him enough for the time he has taken to meet with me, carefully read my work, and provide suggestions for improvement. I have enjoyed working on research projects with him, but most of all, I have enjoyed rocking out with the Messickists. I look forward to our future collaborations as both researchers and musicians.

I would also like to express my deepest gratitude to the members of my dissertation committee, Robert Moll, Craig Wells, and Jennifer Randall. I would like to thank Robert for his helpful feedback, comments, questions, and suggestions. He brought a different perspective to the dissertation that really helped to improve the final document. I am grateful to Craig for his thoughtful input, statistical input, and supportive nature. I could always count on Craig to reassure me that everything was going to go smoothly. And of course I would like to thank Jennifer for her careful reading and editing, as well as her thoughtful comments and questions. Her comments helped me improve as a writer, and I can't thank her enough for that.

In addition to my committee members, I would like to give a very special thank you to Sonya Powers and Pearson for allowing me to use the data for this study. This research would not have been possible without their support and guidance.

I would also like to thank the other faculty members of the Psychometric Methods program, Lisa Keller, Ron Hambleton, and April Zenisky for their constant support. Lisa and Ron’s encouragement and knowledge in statistics and psychometrics have helped make me the researcher I am today, and I can’t thank April enough for collaborating with me on research projects and helping me think about psychometrics in a more applied context.

Of course, I can’t forget my fellow students in the Psychometric Methods program, especially my amazing cohort. To Jerome for dealing with me as an officemate for four years, and letting me bounce ideas off of him and chat with him when I needed a break from work; Rob and Amanda for wonderful morning coffee breaks and trivia Thursdays; Chris for taking the same course schedule as me so we could collaborate on assignments; Minji for her positive energy and for being my singing partner; and Kim for our long talks and for being my fellow ETS intern buddy. I don’t think I would have made it through my doctoral career without the constant encouragement and support of all of you.

Finally, I need to thank James, my rock and sponge, for listening to my conversations about graduate school and psychometrics, even if he didn’t always know what I was talking about. It would have taken me forever to finish if it wasn’t for his pushing me and forcing me to make Google Spreadsheets and timelines. And of course, I would like to thank my family for their tremendous love, support, and understanding throughout these past four years. I know I have made you proud.

ABSTRACT

EVALUATING THE VALIDITY OF ACCOMMODATIONS FOR ENGLISH LEARNERS THROUGH EVIDENCE BASED ON RESPONSE PROCESSES

SEPTEMBER 2013

KATRINA M. CROTTS, B.A., WESTFIELD STATE UNIVERSITY

Ed.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Stephen G. Sireci

English learners (ELs) represent one of the fastest growing student populations in the United States. Given that language can serve as a barrier in EL performance, test accommodations are provided to help level the playing field and allow ELs to better demonstrate their true performance level. Test accommodations on the computer offer the ability to collect new types of data difficult to obtain via paper-and-pencil tests. Specifically, these data can be used as additional sources of validity evidence when examining test accommodations. To date, limited research has examined computer-based accommodations, thus limiting these additional sources of validity evidence. The purpose of this study was to evaluate the validity of computer-based test accommodations on high school History and Math assessments using evidence based on response processes, specifically accommodation use and response time. Two direct linguistic accommodations, non-ELs, two EL groups, and five research questions were investigated in this study.

Accommodation use results indicated significant differences in use across the three student groups, with ELs using accommodations more frequently than non-ELs. However, there were still high percentages of all three groups not accessing any accommodations on individual items. Accommodation use was more common on History than on Math, and decreased as the assessment progressed. Results suggest future research focus on students actually using the accommodations when conducting research on the effectiveness of accommodations.

Response time results showed ELs taking longer to process test items as compared to non-ELs regardless of receiving test accommodations. Receiving accommodations significantly impacted processing time for some of the items on History, but not on Math. Similarly, History showed a relationship between the number of accommodations on test items and response time, but Math did not. These results suggested that the Math content knowledge may have played a larger role in response time than the accommodations. Positive relationships between test performance and response time were found in both subject areas. The most common predictors of both accommodation use and response time across both subject areas were sex, Hispanic status, and socioeconomic status. Implications of the results and suggestions for future research are discussed.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Background

English learners (ELs) refer to students whose first language and/or whose home language is not English, therefore making it difficult for him/her to perform classroom work in English (California Department of Education, 2009; Cawthon, 2010; Pennock-Roman & Rivera, 2011). Sometimes referred to as students with limited English Proficiency (LEPs), students with English as a second language (ESLs), and English language learners (ELLs), ELs comprise a diverse group of students with varying levels of English proficiency, socioeconomic status (SES), expectations of schooling, content knowledge, and immigration status (Hofstetter, 2003; National Council of Teachers in English, 2008). For example, although approximately 80% of ELs across the United States are Spanish speakers, ELs represent over 400 different spoken languages around the country (Pitoniak et al., 2009). The No Child Left Behind (NCLB) Act of 2001 refers to ELs as LEPs, and defines ELs as individuals: (a) age 3 through 21; (b) enrolled or preparing to enroll in an elementary or secondary school; (c) not born in the United States or whose native language is not English; (d) who have difficulties in speaking, reading, writing, or understanding the English language that may deny the individual the ability to meet a state's proficient level of achievement, to successfully achieve in classrooms where English is the language of instruction, or to participate fully in society (NCLB, 2002, Title IX).

ELs represent one of the fastest growing U.S. student populations (Cawthon, 2010), making up nearly 10% of the U.S. student population in 2009-2010 (National

Center for Educational Statistics [NCES], 2012). ELs from a Hispanic background have seen the most growth with the number of Hispanics nearly doubling between 1990 and 2006, suggesting that by 2050, Hispanic students will outnumber those of European descent. In the 2010-2011 school year, approximately two-thirds of all ELs in the United States resided in 5 states including California, Texas, Florida, New York, and Illinois (calculated using NCES Common Core of Data, 2012). Since the implementation of NCLB in 2001, schools have been required to measure and demonstrate the progress of every child, including ELs in grades 3 through 8, and once in high school (NCLB, 2002; U.S. Department of Education, 2005). This requirement of NCLB and the growing population of non-native English speakers throughout the United States have increased the desire to more closely examine EL achievement in comparison to non-ELs on assessments.

Although the United States federal government has provided a definition for ELs through NCLB, EL policy varies across states, individual schools, and school district policies, varying in the definition and identification criteria of ELs, especially when it comes to redesignated ELs (Abedi, 2004; Fry, 2008; Wolf et al., 2008). This variation across states is due to the fact that federal government definitions are used for the purpose of funding allocations, but fail to provide specific operational guidelines (Abedi, Hofstetter, & Lord, 2004; Hofstetter, 2003). Wolf et al. (2008) reviewed different statewide definitions finding that most states define ELs based on students' native language and English language ability in classroom settings. Additionally, they examined how each state identifies ELs, stating the most states administer some type of home language survey and an English proficiency assessment. Some states also gather information from academic achievement tests, informal classroom assessments, teacher observations, checklists, and interviews with the students and parents/guardians (Wolf et al., 2008). Once an EL is identified, test accommodations are considered depending on the English proficiency level and status.

Unlike state EL policy, one area that has been consistent across states is EL performance on standardized assessments. Specifically, research on K-12 standardized assessment performance has found that ELs perform up to one standard deviation (SD) below non-ELs on English and math assessments (Kim & Herman, 2009; Galindo, 2009; Ready & Tindal, 2006), and fail to obtain proficiency and meet adequate yearly progress on statewide assessments (Abedi & Dietal, 2004; Fry, 2008). Similarly, ELs score .72 SDs lower on average than non-ELs on the National Assessment of Educational Progress (NAEP), an assessment administered to students across the United States to provide a common measure of student achievement in 4th, 8th, and 12th grade (Gorman, 2010; National Center for Educational Statistics, 2010 as cited in Cawthon, 2010, p. 2). Different factors have been found to contribute to this achievement gap between ELs and non-ELs on standardized assessments, including time to master academic English, opportunity to learn, linguistic complexity of assessments, reading proficiency levels, and socioeconomic status (e.g., Abedi & Herman, 2010; Abedi, Hofstetter, Baker, & Lord, 2001; Abedi & Lord, 2001; Hakuta, Butler, & Witt, 2000; Ready & Tindal, 2006). Additionally, school-related factors have also contributed to the achievement gap. Specifically, Fry (2008) found that ELs tend to be concentrated in public schools that are typically in central cities and in areas with higher levels of poverty.

ELs tend to perform much lower on reading assessments than in mathematics assessments (Abedi, 2004; Abedi & Dietal, 2004; Sullivan et al., 2005), which is likely due to the linguistic complexity of reading items. For students who speak English as a second language, a test conducted in English could unintentionally result in the test functioning like an English language proficiency test. Because the language barrier causes construct-irrelevant variance, test accommodations are granted to ELs to help "level the playing field" and allow ELs to demonstrate their true ability level without giving them an advantage over students who did not receive the accommodation (Abedi, 2001; Abedi, Courtney, & Leon, 2003a; Abedi et al., 2004; Sireci, Scarpati, & Li, 2005). According to the *Standards for Educational and Psychological Testing*, the main purpose of a test accommodation is "to minimize the impact of test-taker attributes that are not relevant to the construct that is the primary focus of the assessment" (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 1999, p. 101). In the case of ELs, accommodations provide either direct or indirect linguistic support to minimize the language barrier that causes construct-irrelevant variance. More than 75 different types of accommodations are available for ELs (Kieffer, Lesaux, Rivera, & Francis, 2009). Direct linguistic accommodations include glossaries, dictionaries, and read-aloud, whereas indirect linguistic accommodations include different methods of administrating the test such as individual, small group, separate room, and extended time administration (Forte & Faulkner-Bond, 2010). Indirect linguistic accommodations essentially give ELs the opportunity to more adequately process the language in test items, but do not change anything specifically related to the test itself (Pennock-Roman & Rivera, 2011), making indirect test accommodations more common for English language arts assessments. Direct linguistic accommodations tend to be more common for mathematics assessments or assessments that do not focus on language as part of the construct.

Advances in technology have greatly impacted the field of education causing a shift towards technology in the classroom and ultimately computer-based tests (CBTs). For example, the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC) are two consortia currently developing CBTs aligned to the Common Core State Standards in K-12 English language arts and mathematics. In general, CBTs offer advantages over traditional paper-and-pencil tests including: more efficient administration, preference by students, self-selection options for students, improved writing performance, built-in accommodations, immediate results, efficient test development, increased authenticity, and the potential to shift focus from assessment to instruction (Thompson, Johnstone, & Thurlow, 2002). For EL testing, "computers provide an ideal platform for providing flexible options" (Kopriva, 2008, p. 153). Specifically, CBTs offer new attractive innovations and interactions for accommodations that may more effectively remove any construct-irrelevant variance related to linguistic complexity than the accommodations typically provided through paper-and-pencil assessments. For example, students can highlight text, click on graphics, drag objects, self-select font size, magnify graphics, self-select audio, and use pop-up translation (Parshall, Davey, & Pashley, 2000; Thompson, Thurlow, & Moore, 2003). Other technologies include the use of spell-check, speech recognition software, touch screen, calculator, dictionary options, and headphones (Thompson et al., 2003).

Research in both the EL and student with disabilities (SWDs) literature has indicated that computer-based accommodations are more effective than traditional paper-and-pencil accommodations (Abedi, 2009; Abedi, Courtney, & Leon, 2003b; Calhoon, Fuchs, & Hamlett, 2000; Dolan, Hall, Banerjee, Chun, & Strangman, 2005). For example, a common accommodation provided for ELs on a paper-and-pencil assessment are English and bilingual dictionaries. One major issue with this accommodation is that a student must be familiar with dictionaries and need to understand how to use published language tools for the accommodation to be effective (Abedi et al., 2003b). Research has shown the advantage of the computer with respect to the English or bilingual dictionary accommodation. Specifically, Pennock-Roman and Rivera (2011) found in their meta-analysis that the computer pop-up glossary version of the accommodation has been found to be more effective than the paper-and-pencil version because they deliver information more easily and quickly.

In addition to benefiting students with more innovative test accommodations, CBTs can also assist in the validation process increasing the ability to examine different sources of validity evidence more difficult to obtain via paper-and-pencil assessments. The *Standards for Educational and Psychological Testing* (1999) provides five sources for gathering validity evidence based on: (1) test content, (2) response processes, (3) internal structure, (4) relations to other variables, and (5) consequences of testing. Among these five sources, CBTs provide a distinct advantage for gathering evidence based on response processes. Analyzing response processes can "provide evidence concerning the fit between the construct and the detailed nature of performance or response actually engaged in by examinees" (AERA et al., 1999, p. 12). Some methods for gathering this type of evidence include the investigation of student response time or eye tracking. Additional research has also examined cognitive interviews or verbal protocols (Gorin, 2006).

Current research involving validation of test scores with accommodations has focused in three major areas including investigation of the interaction hypothesis, differential boost hypothesis, and measurement comparability (Cho, Lee, & Kingston, 2012). The interaction hypothesis states that the accommodation only improves scores for students who need the accommodation (e.g., ELs or students with disabilities), and does not improve scores for students who do not need the accommodation (Scarpati, Wells, Lewis, & Jirka, 2011; Sireci, Li, & Scarpati, 2003). In relation to ELs, the differential boost hypothesis suggests that ELs will benefit more than non-ELs when provided with the same accommodation (Cho et al., 2012). Lastly, measurement comparability is investigated when an accommodation is functioning appropriately and looks at whether item measurement characteristics function the same for the test administration between accommodated ELs and non-accommodated non-ELs (Cho et al., 2012). Although these three methods investigate the validity of test scores with accommodations in some way, they still fail to examine evidence based on response processes. Understanding response processes of examinees with accommodations could also inform the effectiveness of the accommodations as well. Additionally, gathering evidence based on response processes could potentially help to increase and inform appropriate assignment of accommodations to students. For example, if an accommodation is more effective and valid for an EL with moderate English proficiency, then it might be the case that the accommodation is only valid or effective for moderately proficient ELs and not low proficient ELs.

1.2 Statement of the Problem

In 2004, the National Research Council's Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessments recommended that future research needs to consider other types of validity evidence when examining test accommodations "such as analyses of test content, test-takers' cognitive processes, and criterion-related evidence" (Koenig & Bachman, 2004, p. 7). The authors argued that the current research on test accommodations has been unable to directly address the validity of inferences made on assessments with accommodations. As previously stated, much of the research on the validity of test accommodations has focused on three main hypotheses with a lack of validity research evaluating evidence based on response processes. Although these hypotheses are appealing, they are limited as to how they help to understand how accommodations benefit special populations of students (in this case ELs), support accommodation use, improve test validity, and improve academic instruction (Scarpati et al., 2011).

To obtain evidence based on response processes to examine the validity of test accommodations, accommodation use and response time can be investigated. In relation to accommodation use, previous research on paper-and-pencil assessments have administered surveys to students to determine how often students are accessing an accommodation (e.g., Abedi et al., 2003b). Although self-report measures can be effective, having actual information about student access will better inform whether students are actually using the accommodation, thus indicating whether the accommodation is helping to level the playing field. Evaluation of response time between different subgroups such as ELs with different levels of English proficiency versus non-

ELs, or students receiving an accommodation versus those not receiving an accommodation, is important in ensuring equity on an assessment (Schnipke & Scrams, 2002). CBTs make the response time patterns on individual test items much easier to obtain, especially on accommodated versus non-accommodated items. Understanding how ELs with different background variables respond to test items through the evaluation of accommodation use and response time can help inform the appropriateness of accommodations for ELs of different proficiency levels improving the validity of test scores for examinees with accommodations.

Kieffer et al. (2009) suggested that future studies on test accommodations should consider more innovative methods for accommodating ELs. Currently there is a lack of literature on computer-based accommodations for ELs, which likely goes hand and hand with the minimal research available on evaluating the validity of accommodations based on response processes. To date, only one report (Abedi et al., 2003b) that later became a published article (Abedi, 2009), has examined computer-based accommodations for ELs. As CBTs become more common in standardized assessments, it is essential that more research on computer-based accommodations for ELs be evaluated. More research in this area can help guide the development and usefulness of these computer-based options, options more flexible than those typically found in paper-and-pencil based assessments (Kopriva, 2008).

In addition to focusing on response processes and computer-based accommodations, it is essential that the effectiveness and validity of accommodations also be examined for different student groups (Abedi et al., 2004). Although some studies have examined background variables in relation to test performance for ELs (e.g., Abedi, Lord, Hofstetter, & Baker, 2000; Abedi et al., 2001; Abedi et al., 2003b; Hofstetter, 2003), minimal research has examined these background variables in combination with accommodations. Background variables to consider for ELs include language spoken in the home, English proficiency level, length of time in the United States, and years of schooling in English and/or students' native language (Abedi et al., 2004). In relation to student background variables, Sireci et al. (2003) noted that minimal research has examined accommodations for students in grades 9 to 12. Instead, much of the accommodation literature focuses on grades 4 or 8 rather than high school. Student grade level is an important background variable to consider in addition to other background variables.

The current gaps in the literature in relation to computer-based accommodations, validity of accommodations through evidence based on response processes, and heterogeneity of the EL population make this study an important addition to the EL accommodation literature. Abedi et al. (2004) noted that "new and innovative assessment techniques should be developed and empirically tested to provide approaches and proven effectiveness and validity for all of our students, including English learners" (p. 19), which this study intended to do.

1.3 Purpose Statement

The purpose of this study was to evaluate the validity of accommodations through evidence based on response processes, focusing on accommodation use and response time analysis of ELs as compared to non-ELs on computer-based multiple-choice high school History and Mathematics assessments. Two direct linguistic accommodations including a pop-up glossary tool and sticker paraphrasing tool were examined.

Additionally, this study focused on different levels of English proficiency (mid-proficient and high-proficient) and different student characteristics that could impact these response processes. Specifically, five research questions were addressed including:

1. Do ELs use accommodations significantly more often than non-ELs?
2. What characteristics of ELs and non-ELs predict accommodation use?
3. Do ELs and non-ELs with accommodations take significantly longer to complete items than ELs and non-ELs without accommodations?
4. What characteristics of ELs and non-ELs predict response time on accommodated and non-accommodated test items?
5. What is the relationship between student proficiency, response time, and EL and non-EL accommodation status?

1.4 Significance of the Problem

As more assessments are shifting or being developed to be administered via computer (e.g., PARCC & SBAC), it is essential to examine the validity of the accommodations to ensure fair and equal testing for English learners. Results of this study provide important information to researchers and policy makers to better understand the effectiveness and validity of computer-based test accommodations for ELs. Understanding how EL students use test accommodations and response time patterns can inform test development and can guide how accommodations impact student processes on an assessment. For example, if these new computerized-accommodations require longer time to respond to an item, test developers would need to consider how that could impact overall seat time for students with accommodations. Since seat time for computerized assessments can be expensive, this is essential to consider as many assessments shift to the computer. Evaluating accommodation use and response time could also aid in more accurate interpretations of test scores for ELs. Specifically, response time data may “provide construct validity evidence or illuminate possible sources of construct-irrelevant variance” (Zenisky & Baldwin, 2006). If ELs are taking a long time and repeatedly accessing accommodations for multiple items and still score proficient on the assessment, it could indicate that although the student may be proficient, he/she might be struggling more on the assessment than ELs of the same proficiency level. Essentially, evaluation of response time may be diagnostically useful (Zenisky & Baldwin, 2006). With more high-stakes decisions being made on student test scores including evaluation of teacher effectiveness and student graduation, it is essential that test scores are being interpreted correctly.

Since ELs represent the largest growing student population in the United States, it is especially important to understand how students of differing English proficiency are processing the items on an assessment with accommodations. Essentially, by examining student response patterns through the evaluation of accommodation use and response time, a more complete understanding of what test scores mean for a particular population can be obtained (Gorin, 2006), in this case ELs. Without this validity evidence based on response processes, there is incomplete evidence on interpretations of test scores for ELs when taking an accommodated assessment, which could lead to other inappropriate testing consequences. Ultimately, this study fills a void in the current investigation of validity evidence on test accommodations for ELs by gathering new evidence that can better inform the validity of test scores for ELs.

CHAPTER 2

LITERATURE REVIEW

Research on test accommodations for ELs has increased over the past ten years, but is still very limited. Pennock-Roman and Rivera (2011) stated that more research needs to be conducted on the effectiveness and validity of accommodations. Similarly, Abedi (2004) stated that research on accommodations for ELs is quite “meager,” especially in comparison to research on accommodations for SWDs (Abedi et al., 2004, p. 18). The purpose of this section is to review the current research on the effectiveness and validity of both paper-and-pencil and computer-based accommodations for ELs. Literature on the effectiveness and validity of computer-based accommodations for SWDs will also be examined to inform accommodation research for ELs. In addition to evaluating the effectiveness and validity of accommodations, student background variable impact on overall performance and performance with accommodations will also be discussed. Lastly, research on evidence based on response processes such as accommodation use, response time analysis, and mixture Rasch modeling will be discussed.

2.1 Test Accommodation Research for ELs

Accommodations for ELs are intended to minimize the negative impact or irrelevant language demands on performance, ultimately allowing students to demonstrate their true academic skills and content knowledge (Kieffer et al., 2009). Much of the research on test accommodations has been conducted by Abedi and his colleagues and has focused on both the validity and effectiveness of the accommodations. Abedi and colleagues stated that for an accommodation to be effective, it should minimize the language barrier and enable ELs to demonstrate knowledge in that content area. For an accommodation to be valid it should narrow the performance gap between ELs and non-ELs without altering the construct being measured, that is without affecting the scores of non-ELs (e.g., Abedi et al., 2003a, 2003b; Abedi et al., 2004; Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005). Similarly, Hofstetter (2003) stated that an appropriate accommodation is one that produces an interaction effect (i.e., the interaction hypothesis), meaning the accommodation should improve the performance of ELs, but not change the performance of non-ELs.

To date there have been a series of studies that have either reviewed or conducted meta-analyses on empirical research involving test accommodations for ELs (Abedi et al., 2004; Kieffer et al., 2009; Pennock-Roman & Rivera, 2011; Sireci et al., 2003). The authors have found somewhat similar results throughout the literature regarding the effectiveness and validity of a series of test accommodations including: native language, linguistic modification/simplified English, extra time, customized dictionaries or glossaries, published dictionaries, oral administration, dual-language booklet, and computer-based accommodations. In their review, Sireci et al. (2003) found that small gains for ELs were associated with simplified English and dictionary accommodations. The authors also found that research did not show support for the dual-language booklet accommodation. Abedi et al. (2004) found similar results indicating that research showed support for customized dictionaries and some support for simplified English, finding the accommodations to be effective and valid. The authors also noted that native language translation is only effective if students are given instruction in their native language, and that extra time alone has not shown conclusive evidence of being an effective accommodation.

Both Kieffer et al. (2009) and Pennock-Roman and Rivera (2011) conducted meta-analyses to examine the effectiveness and validity of accommodations. Kieffer et al. (2009) examined 11 empirical studies on EL accommodations from 2001 to July 2006 for a total of 38 effectiveness effect sizes and 30 validity effect sizes. The majority of the studies examined involved students in the 4^{th} or 8^{th} grade taking a mathematics or science assessment, typically involving questions from either the National Assessment of Educational Progress (NAEP) or the Trends in International Mathematics and Science Study (TIMSS). The most common accommodations examined throughout the literature included simplified English, English dictionary or glossary, and bilingual dictionary or glossary. Examining average effect size across different outcomes and grades, results found that in relation to accommodation effectiveness, only the English dictionary and glossary accommodation was found to have a statistically significant and positive average effect size of .018 (p = .001). This accommodation showed no significant moderator effects and reduced the achievement gap by 10-25%. In relation to validity, that is, estimation of increased performance for non-ELs with accommodations, the only significant effect was the Spanish-language translated version accommodation, yielding a negative effect. The result that non-ELs would significantly underperform in relation to ELs on a translated assessment was not surprising.

Pennock-Roman and Rivera (2011) expanded the study by Kieffer et al. (2009) by adding additional studies for a total of 14 empirical studies with 50 different effect sizes from 1990 to 2007. Of the 14 studies, 10 studies overlapped with Kieffer at al. (2009).

Pennock-Roman and Rivera (2011) also expanded the study by categorizing effect sizes by accommodation type together with extended or restricted time, and separating effect sizes for native language accommodations by student English proficiency level and language of instruction. Results indicated that with restricted time, the only significant positive effect was found with the pop-up English glossary accommodation yielding an average effect size of .285 ($p < .05$). For accommodations paired with extended time, the only significant positive effect was found with the English dictionary/glossary accommodation with an average effect of .229 ($p < .05$). When breaking down accommodations by English proficiency level, results indicated that the most effective accommodation for students of low English proficiency was Spanish versions of the test; however, for students with low English proficiency, all effect sizes were below .13 indicating that none of the accommodations were very effective. For students with high intermediate English proficiency, plain English was the most effective accommodation. Overall, the authors noted that it is important to distinguish accommodations with extra time from those with restricted time. They found that the most promising accommodations with extra time included dual-language, bilingual glossary, and the English glossary/dictionary, and that the most promising accommodation with restricted time was the pop-up English glossary.

These four reviews contributed significantly to EL accommodation research. Overall results across these four reviews have showed somewhat similar results, which is likely due to the fact that similar articles were reviewed across all four studies. Overall results have shown support for the simplified English and dictionary/glossary accommodations, which include the pop-up English glossary administered via computer.

2.2 Computer-Based Accommodation Research

Technology offers new opportunities and ways to provide accommodations to students who need them, such as offering the ability to customize a student's test experience through the use of embedded features (National Center on Educational Outcomes [NCEO], 2011). Embedded features are defined as "interactive tools that are part of the test platform and [are] used to customize the assessment for individual test takers" (NCEO, 2011, p. 2). Some states, such as Florida, have already shifted to the computer, and have begun to use these embedded features including features such as font size changes, color contrasting, zooming in and out, and using a screen reader (Beech, 2012). Other features include the use of navigation tools allowing the student to start and stop a reader, move to different parts of the test or reading passage, and change how much text you can see at one time (Beech, 2012). By designing computer-based tests with embedded features in mind, universal design techniques are being implemented that can increase the inclusion of students with disabilities and ELs in testing programs (NCEO, 2011; Thompson et al., 2002). To date, there has been very minimal research on computer-based accommodations for ELs. Fortunately, there has been more research on computer-based accommodations for SWDs, which can help to inform the use of technology for EL accommodations.

2.2.1 Computer-Based Accommodations for ELs

Abedi et al. (2003b) and Abedi (2009) both examined a computer accommodation on 4th and 8th grade assessments using publically released questions from NAEP science and TIMSS. In both studies, the authors examined the effectiveness, validity, and feasibility of a pop-up glossary implemented on the computer, and three traditional accommodations including: a customized English dictionary, extra time, and small group testing. The pop-up glossary, implemented concurrently with extra time, allowed students to scroll over non-content words with a computer mouse to assist their understanding of the test questions. At Grade 8, only the computer accommodation and customized English dictionary were examined.

Controlling for initial differences in English proficiency, results at Grade 4 found the computer-based accommodation to be effective, with ELs scoring significantly higher (p = .005) than ELs taking the standard condition of the test without accommodations. Extra time was also found to be effective (p = .01), but the customized dictionary and small group accommodations were not significant. Results also indicated that the computer accommodation did not affect the construct with non-ELs performing the same both with and without accommodations. Results at Grade 8 were similar, with ELs using the computer accommodation performing significantly higher (p < .01) than ELs under the standard condition. Additionally, non-ELs did not perform significantly different with the accommodation, making it a valid accommodation.

Based on the results of this study, the pop-up glossary computer accommodation was found to be both effective and valid at Grades 4 and 8. The authors also conducted a student survey post-test where ELs indicated that they felt more comfortable with the computer as a form of accommodation as compared to other accommodations. Similarly, students in both grades indicated that they preferred the computer accommodation over other accommodation types. The computer accommodation also offered additional advantages over paper-and-pencil accommodations by presenting items one at a time and being administered in a small setting.

In addition to the two studies above, Bayley, Abedi, and Ewers (2010) presented a study they are planning on doing with computer-based accommodations. Specifically, the authors plan on using both computer-adaptive and non-adaptive computer tests to evaluate student level of English proficiency to obtain information on which accommodation to assign ELs. The computer assessments will use pop-up glossaries, read-aloud versions of the test, and text size adjustment. Although, no official report has come out regarding the results of this study, it shows that more work is beginning to investigate the effectiveness and validity of computer-based accommodations for ELs.

2.2.2 Computer-Based Accommodations for SWDs

Because of the lack of literature on computer-based accommodations for ELs, it is beneficial to also examine the current literature in computer-based accommodations for SWDs. Computer accommodations for SWDs have showed promising results for shifting from paper-and-pencil accommodations to computerized accommodations, showing comparable or slightly higher performance with the computer accommodations, as well as a preference for computer accommodations (e.g., Calhoon et al., 2000; Dolan et al., 2005; Russell, Kavanaugh, Masters, Higgins, & Hoffmann, 2009).

Calhoon et al. (2000) examined computer-based accommodations for SWDs involving students in Grades 9 to 12. Eighty-one students completed a mathematics performance assessment under four different conditions including standard administration, teacher-read text, computer-read text, and computer-read with video. Results found significant differences in scores between the standard administration and each accommodation type ($p < .01$), with effect sizes of .24, .32, and .35 for the teacher-read text, computer-read text, and computer-read with video, respectively. Although no significant differences were found between the three accommodations, 65.5% of the students stated that they preferred the anonymity provided by the computer.

Also examining the computer text-to-speech read-aloud accommodation, Dolan et al. (2005) conducted a 3-week pilot study involving both a paper-and-pencil test (PPT) and CBT with questions from NAEP U.S. History and Civics. Nine 11th and 12th grade students with active Individualized Education Programs (IEPs) were recommended or volunteered for participation in this study. Students received three accommodations on each test including: (1) extended time, (2) direct responses on test booklet or computer, and (3) text-to-speech read-aloud. Overall results indicated that students performed slightly better with the CBT over the PPT; however the results were not statistically significant ($p > .05$). When comparing performance on longer reading passages (more than 100 words) to shorter reading passages (less than 100 words), students performed significantly better when using the CBT as compared to the PPT ($p = .05$). Similarly, results indicated that students identified as "low-average" readers benefitted most from the computer read-aloud accommodation. In an opinion survey, students found the CBT "easy to use and understand" and strongly endorsed the CBT text-to-speech read-aloud accommodation. Approximately 90% of the participants reported using the text-to-speech accommodation on the computer. Participants indicated that the text-to-speech accommodation on the computer allowed them to have more control than with a human reader. This study showed promising results for the computer read-aloud accommodation given its very small sample size.

Extending from the typical read-aloud accommodation, Russell et al. (2009) examined a signing accommodation on the computer for students who are deaf or hardof-hearing. The authors compared a human signer to an avatar signer. The benefit of the computer for these accommodations is that the material is presented in a consistent manner, students can control the size of the video, and they can view the video as often as needed. The advantage of the avatar is that students can select what they want the avatar to look like (e.g., short, tall, brown hair, blonde, etc.), control the background color the avatar signs in front of, activate “lipping” so students can read lips, activate sound that accompanies the signed presentation for students with partial hearing, and switch between American Sign Language (ASL) and signed English. Essentially, students are able to customize their accommodation to make it most effective for them. The study involved 96 middle and high school students taking 8th grade NAEP math items. Results of the study revealed that the majority of students found it easy to perform the test on the computer (77.9%) and liked taking the test on the computer (79.3%). More students found the signing human easier to understand (78.7%) than the avatar (59.7%). However, approximately 53.3% of students reported that the avatar and human were equally effective for communicating test questions. In relation to performance, students did not perform significantly different on individual items when using the human versus avatar.

These three studies involving SWDs show the customization available to test accommodations for students when taking an assessment on the computer. Read-aloud accommodations are available for ELs in addition to SWDs, so the promising results of these studies, even with fairly small sample sizes, suggest promise for computer read-aloud accommodations for ELs. Results indicating SWD preference for computer accommodations are also encouraging and could potentially generalize to ELs as well.

2.3 Student Background Variable Impact on Overall Test and Accommodation Performance

ELs represent a highly diverse group of students, and should not be regarded as a homogenous group with a single defining educational characteristic being use of non-English language (LaCelle-Peterson & Rivera, 1994). The diversity of ELs, represented by different language and background characteristics, can threaten the validity of content-based assessments, as well as the effectiveness and validity of test accommodations (Abedi et al., 2001; Abedi et al., 2004). The *Standards for Educational and Psychological Testing* (1999) also note the importance of considering language background variables for ELs, stating that "it is important to consider language background in developing, selecting, and administering tests and in interpreting test performance (p. 91).

In addition to language background, other individual background variables that could impact student performance and threaten the validity of test scores include: student age, sex, age arrived and length of time in the United States, immigration status, amount of mobility, socioeconomic status (SES), motivation, learning style, and aptitude (Butler & Stevens, 1997). Home, community, and school variables such as home literacy, parent educational background, cultural beliefs, attitudes, expectations, level of parental involvement, ethnic diversity, language use, community attitudes, quality and types of school programs, student opportunity to learn, teacher training and background, and classroom interaction styles could also impact student performance (Butler & Stevens, 1997). In California, a state with more than one-third of the country's ELs, research has indicated that ELs are less likely to have appropriate teachers, curriculum, instruction, assessment, support services, and general learning conditions which is likely to explain some of the reasons for EL underperformance on content-based assessments (Rumberger & Gandara, 2004).

Abedi et al. (2000) conducted a study that resulted in different accommodations achieving different levels of efficacy for different subgroups of students. Specifically, the study looked at NAEP math items and four test accommodations including: (a) modified English, (b) glossary, (c) extra time, and (d) glossary plus extra time. Participants in the study included 946 Grade 8 students from 32 mathematics classrooms in five southern California middle schools. Students were randomly assigned to take the test with one of the four accommodations, or under the standard condition, and were asked to complete a 45-item background questionnaire. Two regression models using the background questionnaire to predict math scores were completed to see if certain accommodations helped some student groups more than others. The full model included the following variables: type of math class, form of accommodation, country of origin, language spoken, television viewing, attitudes towards math, language instruction, and the interactions between these variables. The restricted model included all of the same variables minus the interactions. Results of the full model yielded a R^2 of .281 and the restricted model yielded an R^2 of .251. These two models were significantly different from each other ($p < .01$). These results indicated that the full model had more predictive power explaining a larger amount of variance. Significant predictors included the accommodation main effect, and the interactions between math class and accommodation, and between language of instruction and accommodation. Ultimately, these results indicated that accommodation effectiveness differs according to specific student variables, in this case math class level and language of instruction.

Abedi et al. (2001) used the same sample and data from Abedi et al. (2000). In addition to completing two regression models to look at the interaction between type of accommodation and student background characteristics (the same results as Abedi et al., 2000), the authors also examined the relationship between math and reading scores and student background variables. The authors examined which student characteristic variables best predicted math and reading scores for all students and for ELs only. Using the 45-item background questionnaire taken by each student, results revealed moderate and significant correlations between length of time in the United States and both math ($r(932) = .21$, $p < .001$) and reading scores ($r(932) = .22$, $p < .001$). Moderate and positive relationships were also found between math and reading scores and the following background variables: how long students had studied English and the kind of math the student was taking. Similarly, moderate and negative relationships were also found between math and reading scores and amount of TV students watched in Spanish per day, whether the students spoke a different language, and number of times the student changed schools. All significant correlations were significant at an alpha level of .001.

In addition to correlation analyses, Abedi et al. (2001) also completed regression analyses to predict math and reading scores for all students (ELs and non-ELs), and for ELs only. Results of the regression analysis for all students predicting math score yielded an R^2 of .14, and indicated that for all students the strongest predictor of math performance was the number of years living in the U.S. ($\beta = .20$, $p < .001$), followed by how well the student thought they did in math, how far the student thought they would go in school, and the number of times the student changed schools. The regression analyses predicting math score for EL students yielded an R^2 of .07. Results indicated similar significant predictors; however, for ELs, how far a student thought they would go in school was the strongest predictor, followed by how good a student thought they were in math, and number of years in the United States. Results for reading were very similar with similar predictors as math including: length of time in the U.S., how far a student thought they would go in school, amount of time reading for fun per week, times changed school, and how good the student thought they were in math. When running the analysis with EL students only, the only significant predictor of reading score was how far the student thought they would go in school.

Similar to Abedi et al. (2000) and Abedi et al. (2001), Hofstetter (2003) examined mostly classroom level factors that impact EL performance on NAEP math generally and by test accommodation. Using multilevel modeling, two accommodations were examined including modified English and original Spanish translation. Participants in the study included 849 8^{th} grade students enrolled in 45 math classrooms with 19 teachers in 9 middle schools in a predominately Latino, and low-income area of Southern California. Results revealed no significant interaction between accommodation type and level of math class; however, ELs receiving math instruction in Spanish scored significantly higher using the original Spanish test booklet, suggesting an interaction between language of instruction and type of accommodation. Results of the hierarchical linear modeling (HLM) analyses indicated that the following variables significantly influenced NAEP math performance ($p < .05$): English reading proficiency, Spanish-language instruction, and currently taking an Algebra class. Similarly, both the interaction between the modified English accommodation, English-language instruction and pre-algebra course, and the interaction between the original Spanish accommodation, Spanish-language instruction, and 8^{th} grade math course were significant (for both interactions p = .02). Overall, these results confirm that selected student and classroom variables impacted NAEP math test performance. However, even when variables are controlled, students' level of English reading proficiency, still impact EL and non-EL math performance.

Abedi et al. (2003b) examined two separate regression models for both Grades 4 and 8 to predict performance on English reading scores and on math scores. Predictors for both models were based on a student questionnaire that was completed with the assessment and included: whether the student was born in the U.S., time lived in the U.S., starting grade in the U.S., school resources, how well the student learns math, complaints about math tests, home language before going to school, and language currently spoken in the home. At Grade 4, 14% and 10% of the variance was explained in reading and math score, respectively. For both reading and math, significant predictors included whether the student had attended 1^{st} grade in the U.S., how well the student claimed to be learning math, and student opportunity to learn in math. In reading, the predictor of how often a student complains about math tests was also significant. Amount of variance explained in reading and math score for Grade 8 was similar to Grade 4 at 14% and 13%, respectively. However, for reading, different predictors significantly impacted performance including time in the U.S., how well the student claimed to be learning math, student's use of school resources, and current home language. For math, predictors were similar to Grade 4 and included how well the student claimed to be learning math, student's complaints about math tests, and student's opportunity to learn math. These results suggested that student performance was more often predicted by student's perceptions in math, rather than language factors.

Based on the studies discussed in this section, it is clear that student background variables for ELs impact performance on both overall score and performance with an accommodation. Results have varied for predicting overall score. For example, Abedi et al. (2001) found that amount of time lived in the U.S. was the strongest predictor of math and reading score in Grade 8. How good an EL perceived to be in math and how for an EL thought he/she would go school were also significant predictors. Similarly, Abedi et al. (2003b) found that student performance was often predicted by student perceptions in math, such as how well the student though he/she was learning math, but that amount of time in U.S. was also a significant predictor in both Grades 4 and 8. In relation to accommodation performance, language of instruction and math level appear to be the strongest predictors (Abedi et al., 2000; Hofstetter, 2003). These results show the importance of considering student level variables in relation to performance, especially performance with accommodations.

2.4 Accommodation Use

To date, there has been a lack of research examining how often students are actually using accommodations such as an English dictionary or glossary. Knowing whether ELs are actually using the accommodations can inform both the effectiveness and validity of test accommodations. If students are not using the accommodation, then the accommodation is not providing the linguistic support that an EL might need to level the playing field, therefore impacting the validity of his/her test score. The limited research in this area is likely due to limited research on computer-based accommodations, where it is easier to obtain information on how often students might use specific accommodations.

In their study examining computer testing as a form of accommodation for ELs, Abedi et al. (2003b) also examined accommodation use. The authors examined the effectiveness, validity, and feasibility of multiple accommodations including: a pop-up glossary implemented on the computer, a customized English dictionary, extra time, and small group testing. During this study, the authors examined how often students used the customized English dictionary (a paper-based accommodation) using a follow-up accommodation questionnaire. Results of the survey revealed that very few students indicated that they used the customized English dictionary. During the assessment, ELs were administered a sample word to look up using the accommodation. In Grade 8, 140 of 204 (~69%) students had marked the sample word. In addition to the sample word, only a maximum of 4 students marked any given word on the page of definitions. The limited number of students using the dictionary in Grade 8 may have been related to the fact that only 14% of all students stated they had used a dictionary in class before, and 15% of all students stated that they would use an English dictionary to help them understand math problems. Similarly, only around 13% of Grade 8 students indicated that they would prefer the customized English dictionary over other accommodations. In Grade 4, 146 of 176 (~83%) of EL students marked the sample word, and a maximum of 8 students marked any given word on the page of definitions. Observations by the test administrator and the survey suggested that limited accommodation use may have also been related to the fact that if students did not find a word defined in the customized dictionary within the first few attempts of using the accommodation, they may have stopped using the accommodation.

In addition to the customized English dictionary, Abedi et al. (2003b) also examined how many words ELs looked up on average using the computerized pop-up glossary accommodation. Because this accommodation was administered via the computer, the computer was able to record how many words each examinee glossed over on average. Results for Grade 4 students (n = 35), indicated that ELs looked up 17.5 words on average (SD = 10.3), compared to 18.9 words on average (SD = 9.5) for non-ELs (n = 44). At Grade 8, ELs glossed over twice as many words as non-ELs. Grade 8 ELs (n = 84) looked up 26.0 words on average (SD = 14.9) compared to 15.7 words on average (SD = 10.0) for non-ELs (n = 68). This difference in average number of words glossed over between ELs and non-ELs at Grade 8 was statistically significant ($p < .001$). These results suggest that ELs at Grade 8 may more effectively use the accommodation than ELs at Grade 4, where ELs used the accommodation less than non-ELs. In comparison to the results on accommodation use for the customized dictionary, it appears that students using the computer accommodation looked up more words than those using the paper-and-pencil accommodation.

2.5 Response Time Analysis

Computer technology has not only offered the ability to customize testing for examinees through the use of new test accommodations, but it has also offered the ability to collect additional information on student response processes on an assessment. Specifically, response times on individual test items and full tests for all students can be collected. Response time analysis is not new to the testing field and has existed, especially in the cognitive psychology literature, since the mid-1950's (Schnipke & Scrams, 2002). However, the use of computer technology has resulted in increased availability of response time information allowing for the integration of response time into routine test development and validation practices (Zenisky & Baldwin, 2006). Much of the research in response time has focused on areas such as scoring, speed-accuracy relationships, speededness, pacing, setting time limits, and subgroup differences (Schnipke & Scrams, 2002). Of these research areas, the most applicable to this study include research on speed-accuracy relationships and subgroup differences.

Research on speed-accuracy relationships have examined whether a student's performance level impacts how long they spend on test items. Research in this area has found that high performing examinees have different response time patterns as compared to low performing examinees. For example, research has shown that students with high performance tend to take longer on test items as compared to students with lower performance (Chang, Plake, & Ferdous, 2005). In general, research has indicated that examinees spend more time on items they answer incorrectly than items answered correctly (Chang et al., 2005; Hornke, 2000, 2005). In relation to performance level, Chang et al. (2005) noted that higher performing examinees tend to spend more time on items they answer incorrectly compared to items answered correctly, whereas lower performing examinees spend roughly the same amount of time on items regardless of whether the item was answered correctly or incorrectly. Additionally, higher performing examinees are more likely to distribute their time effectively throughout the entire assessment, whereas lower performing examinees are more likely to take longer in the beginning of the assessment and then rush and guess on items towards the end of the assessment (Giraud & Smith, 2005).

In relation to this current study, research surrounding subgroup differences in response time, specifically differences between ELs and non-ELs is of upmost interest. According to Schnipke and Scrams (2002), “examining subgroup differences in response time is not only possible, but necessary to ensure equity” (p. 260). Although differences in response time rates may not be directly related to differences in item-level performance, if the assessment is timed, it could result in the test being speeded for some subgroups, which could ultimately impact overall student performance (Schnipke & Pashley, 1997). To date, much of the literature surrounding this topic has been presented at national conferences such as the National Council on Measurement in Education, but little research has been published in journal articles. Additionally, this literature has mostly focused on sex and ethnic differences, rather than linguistic differences. Results across much of the literature are mixed with some studies indicating that sex and ethnicity are not significant predictors of response time (e.g., Bergstrom, Gershon, & Lunz, 1994; Parshall, Mittelholtz, & Miller, 1994; Schnipke, 1995). However, across other studies, small differences have been found (e.g., Llabre & Froman, 1987; O’Neill & Powers, 1993; Schnipke & Pashley, 1997). Schnipke and Scrams (2002) suggested that these differences might be small because they are being masked by other predictors such as item difficulty or word count. Additionally, even though results have been mixed, it is important to closely examine subgroup timing differences to ensure that there are no disadvantages for those subgroups when taking a timed assessment (O’Neill & Powers, 1993).

Llabre and Froman (1987) conducted an early study using microcomputers to examine Hispanic and White examinee time allocation patterns for the purposes of explaining differential test performance. The study involved 28 White students and 38 Hispanic students enrolled in a beginning algebra course at Miami-Dade community college. Taking an untimed 16 item multiple-choice inference subtest of the California Test of Mental Maturity, results indicated significant main effects for ethnic group and item on response time. Similarly, the interaction was also significant ($p < .001$). Results revealed that on average, Hispanic students scored 1 point less than White students and took 6 minutes longer to complete the assessment. The authors speculated that had there been a 10 minute time limit on the assessment, that Hispanic students would have performed 6 points lower than White students. In a correlation analysis between mean item time and item difficulty, results found a stronger relationship for White students indicating that White students allocated their time according to difficulty of the item to a greater extent than Hispanic students. Since item difficulty was the same for all items except for one, the authors suggested that the reason for Hispanic students needing more time might be due to the need to translate items from English to Spanish, thus resulting in more processing time.

Schnipke and Pashley (1997) examined the distributions of response times for both ELs and non-ELs taking a non-adaptive high-stakes computer assessment with 25 items. The sample contained 6,306 non-ELs and 462 ELs. Using survival analysis with a covariate of test score because of its influence on response time, results indicated that test score was a significant predictor of response time for all items. Similarly, level of English fluency was a significant predictor for about half the items. Results revealed that ELs responded slower on average to the items on the first half of the test as compared to non-ELs, but responded faster on average to the items on the second half of the test. Additional analyses confirmed this faster pace as rapid-guessing behaviors, which the authors noted could be linked to the items being slightly more difficult on the second half of the test. Overall this study revealed that for many of the items, significant differences between response time across ELs and non-ELs were found, which supports the need to examine response time differences in different subgroups.

Zenisky and Baldwin (2006) conducted a study where they examined EL and non-EL adults enrolled in an adult basic education program in a Northeastern state. Data included 3,284 students completing a 40-item math test, and 3,254 students completing a 40-item reading test. For math there were four levels of item difficulty including low, medium, high, and advanced, and for reading, there were only three levels of item difficulty. As part of their study, the authors analyzed response time and overall performance, as well as the relationship between subgroup membership, cognitive dimension of the items, item complexity, and response time. Results yielded significant differences between ELs and non-ELs for all levels but the lowest test difficulty levels in both math and reading. In math, ELs spent approximately 8-10 minutes longer on the test than non-ELs, and in reading ELs spent approximately 6-7 minutes longer. In relation to overall performance, ELs performed significantly higher in the middle two levels of math. In reading, non-ELs performed significantly higher than ELs at the highest level. As expected, all students in both math and reading took longer to complete more difficult test items, with ELs taking longer than non-ELs.

Abedi et al. (2003b) took their study further than other studies and examined response time differences for ELs and non-ELs with a pop-up glossary accommodation. Specifically, the authors examined how long (in seconds) ELs spent using the pop-up glossary. During the assessment, the computer recorded how long examinee spent using the glossary. Results for Grade 4 students indicated that ELs (n = 35) spent an average of 65.6 seconds (SD = 55.9) on the pop-up glossary items. This was compared to 68.7 seconds (SD = 52.3) for non-ELs (n = 44). The difference in average time between ELs and non-ELs was not statistically significant. At Grade 8, ELs spent nearly three times as much time using the glossary as non-ELs. Specifically, ELs (n = 84) spent an average of 188.6 seconds (SD = 206.3) on pop-up glossary items compared to 65.9 seconds (SD = 72.3) for non-ELs (n = 68). The difference in average time between ELs and non-ELs was statistically significant ($p < .001$). These results suggest that at Grade 8, ELs tend to spend longer on accommodated items, than ELs using an accommodation at Grade 4.

In the student with disabilities literature with accommodations, Russell et al. (2009) examined differences in length of time to complete a test with a human avatar versus an avatar. Focusing on full test response time rather than item-level response time, results revealed that for Form 1, students taking the test with the human avatar took approximately 9.89 minutes on average (SD = 4.14) on the assessment. Similarly, students taking the test with the avatar took approximately 10.23 minutes on average (SD = 10.23). Form 2 revealed similar results with students taking 8.40 minutes on average (SD = 4.54) with the human avatar and 9.32 minutes on average (SD = 4.76) with the avatar. Overall results indicated that neither Form 1 nor Form 2 had significant differences in time between accommodation type received.

Research on response time has found that both student performance level and subgroup differences can impact response time on assessments and individual test items. Similarly, research has suggested differences in response time levels when students receive an accommodation and when they do not. Additionally, grade level with accommodations may also impact the amount of time students spend when using an item with an accommodation (see Abedi et al. 2003b). With similar accommodations at the same grade level, results have revealed no significant differences in average response time (Russell et al., 2009).

2.6 Mixture Rasch Modeling

Mixture Rasch modeling (MRM) is a method that combines both the Rasch model and latent class analysis (LCA). The Rasch model is an item response theory model that models the probability of a dichotomous (correct or incorrect) item response as a function of person and item parameters, specifically student ability and item difficulty (Wright & Masters, 1982). LCA is method used to identify subpopulations not distinguishable on the basis of observed features, thus making them latent classes (Yang, Shaftel, Glassnapp, & Poggio, 2005). LCA is similar to factor analysis in that both models posit an underlying latent variable measured by observed variables; however, in LCA the latent variable is categorical (Collins & Lanza, 2010). MRM is a method that uses the Rasch model to describe the response behavior of examinees within a latent class obtained through LCA, meaning that different sets of item parameters are obtained for the different latent classes (Rost, 1990). Much of the research involving MRM and test accommodations has been in the student with disability literature (e.g., Cho et al., 2012; Cohen, Gregg, & Deng, 2005; Scarpati et al., 2011), rather than the EL literature. This area of research has focused on using MRM in combination with differential item functioning (DIF) to attempt to explain performance differences on assessments between different subgroups.

Cohen et al. (2005) investigated the role of the extended time accommodation and math content knowledge on secondary students' performance outcomes. The authors focused on the extended time accommodation due to its prevalence as an accommodation in statewide assessments. The authors conducted two studies. The first study identified DIF items followed by mixture IRT to define groups of students whose response patterns were consistent with the pattern of accommodation-related DIF. The study involved 1,250 students with disabilities (SWDs) with an extended time accommodation, and 1,250 non-accommodated students without disabilities (SWODs) taking the Florida statewide assessment with 29 multiple-choice items. All students were randomly sampled from a larger sample of students. Results of Study 1 indicated that 22 items had some amount of accommodation related DIF with 13 items being easier for the accommodated group, and 9 items being easier for the non-accommodated group. Results of the two-group mixture Rasch model indicated that 62% of the students in the sample responded in a manner consistent with students receiving an accommodation (Class 1), and 38% responded in a manner consistent with students receiving no accommodation (Class 2). Accommodation and class membership were moderately correlated at .35. Since only 67% of students were assigned to the same latent class as their accommodation would suggest, results indicated that students' accommodation status is not a sufficiently useful explanation variable for determining cause of DIF performance.

For Study 2, Cohen et al. (2005) wanted to know whether there was some other way of identifying latent classes of students that performed differentially on the 29 multiple-choice items. A second sample the same size as Study 1 was drawn from the full sample for this study. Results of the MRM suggested a three class solution. Of the three classes, Class 3 had the highest mean score followed by Class 1, then Class 2. Results indicated that different items were disproportionately easier for each class. Results of an ANOVA revealed a significant main effect for accommodation status, latent class, and the interaction between the two on mean math score ($p < .05$). In relation to student background characteristics, Class 2 had the highest percent of accommodated students, followed by Class 1 and Class 3, respectively. With regards to ethnicity, Class 2 had the lowest percentage of White students and highest percentage of African American and Hispanic students. Both accommodation and sex were not associated with class membership. Ultimately, these results suggest that group differences on item-level performance were associated with differential difficulty of math content rather than accommodation, and that the use of accommodation status contributes little to understanding why students differ in test performance. This study provided an alternative method for investigating the influence of accommodations on test scores through the use of mixture modeling.

Similar to the study conducted by Cohen et al. (2005), Scarpati et al. (2011) compared item difficulty between SWDs with accommodations (i.e., use of a calculator or a presentation accommodation) and SWODs without accommodations to determine whether the accommodation was primarily responsible for any observed differences. The study involved 73,000 students, 12,268 being SWDs with accommodations who took an 8^{th} grade math assessment involving 34 dichotomous items. Results focusing on the calculator accommodation revealed that 14 of 34 items exhibited DIF with 8 items being easier for the accommodated group. Results of the two-group MRM indicated that 47% of students exhibited item responses consistent with non-accommodated students. Of these students, 19% were those students who had actually received the calculator accommodation. Similar results were found with the presentation accommodation with 9 or 34 items showing DIF, and 56% of students exhibiting item responses consistent with non-accommodated students. However, of that percent, 64% had actually received a presentation accommodation even though they responded more similar to those without. With both sets of results, students with accommodations whose responses were consistent with the non-accommodated group performed nearly 1 SD larger than their counterparts in Class 2. This study showed the interaction between student ability and receiving an accommodation. Specifically, when ability levels vary, the influence of the accommodation varies as well, and that differences in performance are most associated with math skill level.

In a more recent study, Cho et al. (2012) investigated accommodation validity from the perspective of understanding the relationship between DIF, item types and item features, and students' accommodation status and content knowledge. This study involved 1,770 SWDs with accommodations (i.e., frequent breaks, separate quiet setting, or read aloud), and 49,821 SWODs without accommodations in grades 3-8 taking statewide math assessment. The authors implemented item analysis, DIF analysis, and mixture modeling analysis. The two-class mixture model analysis constrained item parameter estimates so that only non-DIF items were equal between the focal and reference groups. The goal of this analysis was to see if accommodation status was the primary factor that contributed to the observed DIF. Results yielded a total of 101 of 470 items flagged as functioning differently between the focal and reference groups with a mix of uniform (62 items) and non-uniform (36 items) DIF, and 25 items favoring the focal group, and 48 items favoring the reference group. Results of the mixture model analysis suggested that not all students in the focal and reference groups were consistently advantaged or disadvantaged by the DIF items. Specifically, the proportion of SWDs in Class 1 ranged from .61-.78 depending on the grade, and from .22-.39 in Class 2. When examining differences between the accommodated SWDs in the two classes, results indicated no significant differences in sex, disability category, ethnicity, nor latent ability. SWDs classified into Class 2 had significantly higher math proficiency than their counterparts in Class 1 in Grades 3 and 5. This study showed how mixture modeling can be used to understand the interaction between student accommodation status and academic ability with regard to DIF. In this study, no consistent interaction was found.

Although to date, much of the literature involving MRM is related to accommodations for SWDs, the literature stills shows how this method can be applied to ELs with accommodations. Results across the research has been fairly consistent in finding that group differences on item-level performance tend to be more associated with math ability level rather than accommodation (Cohen et al., 2005; Scarpati et al., 2011). However, Cohen et al. (2005) also showed how MRM can be used to also examine other student background characteristics such as sex and ethnicity. Since the authors found that accommodation status contributes only small amount of understanding for group differences, it could be that other background characteristics could contribute to group differences. In relation to ELs, this could include student language ability, language spoken in the home, or amount of time living in the United States.

2.7 Summary of Literature Review

This literature review discussed some of the current literature on test accommodations for ELs, and discussed different analyses that can be examined with the use of computer accommodations. In relation to accommodations for ELs, the literature has showed support for simplified English and dictionary/glossary accommodations. Only two studies (Abedi et al., 2003b; Abedi, 2009) that used the same data, examined a computer accommodation of a pop-up English glossary. Results are promising for this accommodation, finding it to be both effective and valid. Additionally, students indicated that they also preferred the computer accommodation over paper-and-pencil based accommodations. Because of the limited availability of research on computer accommodations for ELs, the SWD literature was also examined. Studies in this area have also shown support and student preference for computer accommodations (Calhoon et al., 2000; Dolan et al., 2005; Russell et al., 2009). Student background variables have also been found to impact EL performance with and without accommodations. Across the literature, studies have indicated that length of time living in the U.S., language of instruction, math level, and student perception of math ability all impact performance. These results suggest this importance of considering background variables when examining accommodations for ELs.

In addition to accommodation literature, this literature review also discussed different approaches and methods to gather evidence based on response processes to examine the validity of test scores with computer accommodations. Specifically, research on accommodation use was discussed revealing that students using the pop-up glossary on the computer tended to look up more words than students using the customized dictionary administered via paper-and-pencil (Abedi et al., 2003b). Research on response time has indicated subgroup differences on response time (Llabre & Froman, 1987; Schnipke & Pashley, 1997; Zenisky & Baldwin, 2006), but to date, only one study in the EL literature looked at differences in response time when receiving an accommodation (Abedi et al., 2003b). Results indicated differences in response time depending on student grade level, indicating that at Grade 8, ELs tend to spend longer on accommodated items than non-ELs. Examining response time on accommodated test items is important for understanding how students are processing test items.

The last area that was discussed in the literature review was the use of MRM. Currently, research being conducted using this method of analysis with accommodations has been within the SWD literature. Results across the literature have suggested that ability contributes more to group differences rather than receiving an accommodation (Cho et al., 2012; Cohen et al., 2005; Scarpati et al., 2011). Literature on MRM also suggests that student background variables could also be contributing to differences in response patterns when receiving an accommodation (see Cohen et al., 2005). Initially, the current study intended to use the MRM method, however, given the results of research conducted with SWDs, it was decided that the focus on the current study would be on the relationship between student performance and response time patterns, rather than the use of MRM analysis.

At present, there are no empirical studies examining EL and non-EL response time and patterns with and without accommodations, especially while focusing on the heterogeneity of the EL population. Results of this study will fill the void in the literature on computer accommodations for ELs. Because the pop-up glossary accommodation was found to be effective in previous studies (Abedi et al., 2003b; Abedi, 2009), this study will not only fill the gap, but will also build on the current literature by also examining pop-up glossary accommodations. Additionally, this study will take on new approaches for evaluating the effectiveness and validity of test accommodations by extending the research from evaluating the interaction hypothesis, to evaluating how certain student groups are actually processing the test items with and without test accommodations.

CHAPTER 3

METHODOLOGY

3.1 Overview

The main goal of this study was to provide evidence based on response processes to evaluate the validity and effectiveness of test accommodations for English learners (ELs). To achieve this goal, this study used data collected from an empirical computer-based accommodation study focusing on student response time and how often students are accessing the accommodations. Specifically, the following series of research questions were addressed:

1. Do ELs use accommodations significantly more often than non-ELs?
2. What characteristics of ELs and non-ELs predict accommodation use?
3. Do ELs and non-ELs with accommodations take significantly longer to complete items than ELs and non-ELs without accommodations?
4. What characteristics of ELs and non-ELs predict response time on accommodated and non-accommodated test items?
5. What is the relationship between student proficiency, response time, and EL and non-EL accommodation status?

This chapter begins with an in depth description of the sample, the assessment, and the test accommodations used throughout this study. This chapter also describes how each series of research questions were addressed with different analyses.

3.2 Sample

The data for this study were from a statewide study conducted to examine the effectiveness of computer accommodations for ELs on both History and Math assessments. For both assessments, across two test forms (described in the following section), the sample consisted of ELs and non-ELs in high school who were currently enrolled in either a History or Math course, resulting in 2,565 and 2,192 students taking the History and Math assessments, respectively. ELs were randomly sampled across school systems with large numbers of ELs, and non-ELs were randomly sampled across the state. This sampling strategy resulted in three groups of students including mid-proficient ELs, high-proficient ELs, and non-ELs. Mid-proficient ELs were those ELs with mid-proficiency in reading living in the United States for three or fewer years. Similarly, high-proficient ELs were those ELs with high-proficiency in reading with four or fewer years in United States schools. The reason for categorizing ELs based on reading proficiency and length of time in the United States was consistent with Abedi et al. 2004, who stated that the most commonly reported criteria across states for categorizing ELs includes both formal English language proficiency assessments and time spent in the United States or English-speaking schools. It is important to categorize ELs based on English reading proficiency because accommodations that are appropriate for EL students with high levels of English proficiency might not be relevant for EL students with low levels of English proficiency (Abedi et al., 2004). The demographic information for these three groups on each assessment across two test forms can be found in Tables 3.1 and 3.2 for History and Math, respectively. After examining the demographic information, it was decided the students with disabilities would be removed from the analyses since those students might benefit from the use of the accommodations which could impact any potential results.

3.3 Assessment

The History and Math assessments used in this study included statewide linguistically accommodated test items and non-accommodated test items. The History assessment contained 25 multiple-choice items, and the Math assessment contained 23 multiple-choice items, and two short answer items. The two test forms within each subject area contained the same items; however, different items were accommodated on each form. For example, if item 1 was accommodated on Form 1, it was not accommodated on Form 2, and vice versa (see Figure 3.1). Additionally, there were some items across each form that were not accommodated because they did not require any word clarification. On the History assessment a total of 23 items were accommodated across the two forms, and on the Math assessment a total of 19 items were accommodated across the two forms.

The assessments were randomly assigned to students via computerized adaptive spiraling to get a balance of ELs and non-ELs receiving the two different test forms. Upon beginning the assessment, students were provided with instructional training containing sample items to ensure student understanding with the accommodation system and its functionality.

3.4 Computer-Based Accommodations

Accommodations used throughout the assessment were all direct linguistic accommodations, providing clarification for students on words and language structures that could be unfamiliar to students of lower English proficiency. The accommodations in this assessment provided clarification through the use of definitions, synonyms, paraphrase, pictures, and animations. Specifically, two different tools provided clarification for students including a pop-up glossary tool and a sticker paraphrasing tool.

The pop-up glossary tool provided an accommodation with the use of the mouse cursor. Students could click on pre-identified (underlined) words or phrases unassociated with the test content that could still contribute to construct-irrelevant variance for students with lower English proficiency. Upon selecting the words or phrases, a window would show clarification of through the use of definitions, synonyms, pictures, or animations. Figure 3.2 shows an example of the pop-up glossary tool with pre-identified words underlined in blue. If the student clicked on the words, the windows would appear as shown in Figure 3.2.

The sticker paraphrasing tool provided an accommodation on larger portions of text through the use of paraphrasing. Students select an icon that then unrolls like a sticker on top of the original text. Stickers are essentially used in areas where a large number of pop-ups would be due to complexity in the text, and therefore tended to appear less often than pop-up items. Figure 3.3 shows an example of this accommodation. The left side of the figure shows how the question would look prior to selecting the accommodation, and the ride side of the figure shows how the question would appear with the accommodation.

It is also important to note that one item could have both accommodations. For example, if the item contained a lot of text, the sticker paraphrasing tool would be used, but there still could be underlined words even within the newly paraphrased text. The pop-up glossary tool appeared in items more often than the sticker paraphrasing tool.

3.5 Data Analyses

Prior to running the analyses, the test forms were restructured for each subject area, creating an accommodated form (Form 1A) and a non-accommodated form (Form 2A) to be used in some of the analyses. It is possible to restructure the test forms since test items were the same across both test forms, and students were randomly assigned to each form meaning that the students are roughly equivalent across the Forms 1 and 2. Demographic equivalence across the two test forms was shown for History and Math in Tables 3.1 and 3.2, respectively.

Test forms were restructured based on the number of accommodated items within a subject area. For History, there were a total of 23 accommodated test items across Forms 1 and 2 (2 items did not include language that needed an accommodation). This resulted in Form 1A having a total of 23 test items. Similarly, for Math, there were a total of 19 accommodated test items (6 items did not include language that needed an accommodation) resulting in Form 1A having 19 items. Form 2A across both subject areas contained all 25 items. Figure 3.4 shows an example of the new test forms used for analyses.

3.5.1 Accommodation Use Analyses

Student accommodation use can be obtained through the use of a computerized assessment. For each accommodated item, it was possible to have multiple accommodations. For example, two words could be underlined for the pop-up glossary, and sticker paraphrasing could be available, resulting in three accommodations for that one item. The number of accommodations on a single item ranged from 1 to 11 and 1 to 8 for History and Math, respectively. Items could have the pop-up glossary tool, the sticker paraphrasing tool, or both. Additionally, it was possible for pop-up glossary words to be nested within the sticker paraphrasing tool. These differences in the number of accommodations across items results in different amounts of accommodation use per item. Students could access single accommodations as many times as they see fit, however, across most items, students most often accessed the item once or twice, most likely because the accommodation stayed open upon selecting it. Therefore, in evaluating accommodation use, accommodation use was coded as 0 or 1 to allow for examination of students accessing the accommodation at least once.

3.5.1.1 One-Way ANOVA

To investigate overall differences in accommodation use between the non-EL and EL student groups (Research Question 1), a one-way analysis of variance (ANOVA) was conducted. Specifically, the three student groups were the independent variable, and total accommodation use was the dependent variable. Total accommodation use varied by subject area and across original test forms. Specifically, History had a total of 53 and 60 accommodations on Forms 1 and 2, respectively, resulting in a total of 113 accommodations available. Similarly, Math had a total of 45 and 42 accommodations available on Forms 1 and 2, respectively, resulting in a total of 87 accommodations available across the two test forms. Examining total accommodation use differences between student groups will give a sense of which student groups were accessing the accommodations more frequently on the total assessment.

3.5.1.2 Chi-Square Test

To further examine whether ELs use accommodations significantly more often than non-ELs (Research Question 1), chi-square analyses were conducted separately for each accommodated item in each subject area. Because each test item has different numbers of accommodations available, that information was taken into account when performing each chi-square. For example, if there were three accommodations available on an item, the contingency table would look like Figure 3.5 to investigate the differences in accommodation use across the three student groups. As seen in Figure 3.5, examinees in the 1 accommodation use column used one accommodation on the item at least once, and examinees in the 3 accommodation use column used all three accommodations on the item at least once.

The chi-square test for independence compares counts of categorical responses between two groups. To conduct the chi-square test, a contingency table (see Figure 3.5) was first created to obtain counts. The chi-square statistic was then calculated using Equation 3.1,

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \tag{3.1}$$

where f_o is the set of observed frequencies, and f_e is the set of expected frequencies (Gravetter & Wallnau, 2002). Expected frequencies were calculated by multiplying the marginal frequencies for the row and column of the desired cell and then dividing by the total number of observations:

$$f_e = \frac{RowTotal * ColumnTotal}{N} \tag{3.2}$$

The chi-square test across each item allowed for investigation of significant differences in accommodation use across the three student groups. Cramer’s V was used to evaluate the effect size of each chi-square test and is calculated as follows:

$$\phi_c = \sqrt{\frac{\chi^2}{N(k-1)}} \tag{3.3}$$

where N is the total number of observations and k is the number of rows or columns in the contingency table. Cramer's V varies between 0 and 1, with 0 indicating no association, and 1 indicating a strong association between student group and accommodation use. Cramer's V values less than .20 have a weak or negligible association. Values between .20 and .40 are considered a moderate association, and values greater than .40 have a relatively strong to very strong association (Rea & Parker, 1992).

To help in interpretation of these results, patterns were investigated between the size of Cramer's V and number of accommodations, item difficulty, and item location. Specifically, graphs were created to evaluate whether students were accessing accommodations more at the beginning of the assessment than towards the end of the assessment. Additionally, frequencies of accommodation use on each item across the three student groups in both subject areas were shown in graphical form to aid in the understanding of accommodation use differences at the item level.

3.5.1.3 Poisson Regression

With accommodation use the focus is on counts, or number of times a student clicks on an accommodation. With count data, it is inappropriate to use ordinary least squares (OLS) regression due to non-normal distribution of residuals and heteroscedasticity of residuals thus resulting in prediction bias. Additionally, since counts cannot go below zero, predicted scores using OLS regression could be out of range, and regression coefficients may be biased and inconsistent. Lastly, standard errors using OLS regression could be underestimated and thus inflate *t*-tests for individual regression coefficients (Cohen, Cohen, West, & Aiken, 2003). With the issues surrounding OLS regression, Poisson regression was used instead to address which characteristics of ELs and non-ELs predict accommodation use (Research Question 2).

To understand Poisson regression, it is first helpful to understand the Poisson probability distribution. In relation to accommodation use, the Poisson distribution shows the probability that an examinee will use a specific number of accommodations given the length of the test. Each distribution has a rate parameter (the average number of events expected in that time period), μ, that differs across distributions (Cohen et al., 2003). One important property of the Poisson distribution is that the mean and variance of the distribution are equal, meaning the variance is completely determined by the mean of the distribution (Cohen et al., 2003). When this property or assumption is not met such that the variance is greater than the mean, the data are said to be overdispersed resulting in the standard errors of regression coefficients to be too small, and resulting in significant predictors being overestimated. Therefore, it is important to check the dispersion parameter, φ, which equals 1 if the assumption is met. If this assumption is not met, the negative binomial regression model will be used instead. Both the Poisson regression and negative binomial regression model predicts the number of events ($\hat{\mu}$) from values on a set of predictors X_1, X_2,…X_k as shown below:

$$Ln(\hat{\mu}) = B_1 X_1 + B_2 X_2 + ... + B_k X_k \qquad (3.4)$$

With Poisson regression, the expected number of times students used the accommodations on the accommodated test items from Form 1A for each subject area was predicted. Predictors included: sex, ethnicity, SES, at-risk status (students at risk of dropping out of school under state-mandated academic criteria), and statewide ELA score for non-ELs, with the additions of number of years in the U.S. and English proficiency test score for ELs. Accommodation use will be the dependent variable. Three Poisson regressions were analyzed for each of the student groups across both subject tests. For each of these models, the assumption in relation to the conditional mean and variance, as well as model fit were investigated prior to interpreting the results to determine whether negative binomial regressions should be used instead.

3.5.2 Response Time Analyses

Similar to accommodation use information, because the assessment was administered via the computer, each item contained response time information (in seconds) that can be used for further analysis. Verbic and Tomic (2009) define response time as "the time elapsed between presenting the question on the computer screen and the response to that question" (p. 3). Because the distributions of response times are positively skewed, the natural logarithm of response time (Ln(RT)) is used throughout the literature to make the distribution more normal (Scrams & Schnipke, 1999; Thissen, 1983; van der Linden, 2006). The last three research questions involved the use of response time with each analysis described in more detail below.

3.5.2.1 Descriptive Analysis

Descriptive analyses were conducted for all students, and for separate student groups (mid-proficient ELs, high-proficient ELs, & non-ELs) on both the original test Forms 1 and 2, and on test Forms 1A and 2A (accommodated & non-accommodated forms) for each subject area. For the original test forms, median total response time was calculated (in minutes). Similarly, the mean and standard deviation of Ln(RT) were also calculated. The same information was obtained for Forms 1A and 2A; however, response time was calculated at the item-level (in seconds). Specifically, the median of median response time per item, and the mean and standard deviation of Ln(RT) per item were calculated. To interpret Ln(RT), response time was converted back to minutes or seconds by taking the base of the natural logarithm, exp(Ln(RT)).

3.5.2.2 Two-Way ANOVA

To examine whether ELs and non-ELs take significantly longer to complete items than ELs and non-ELs without accommodation (Research Question 3), a two-way ANOVA was completed for each subject area. The two independent variables were test form (Forms 1A & 2A) and student group (mid-proficient ELs, high-proficient ELs, & non-ELs). For the purposes of this analysis, the number of items determined the sample size, and the dependent variable was the mean of Ln(RT) for each item. With the two-way ANOVA, statistical differences on response time between groups, between accommodated and non-accommodated items (Forms 1A & 2A), and the interaction between the two were examined. Because Type I error (i.e., false positives or rejecting the null hypothesis when it is true) can occur when performing multiple analyses, family wise error was controlled for using Fisher LSD (because number of groups = 3) when looking at differences between groups and at the interactions.

3.5.2.3 Standardized Mean Difference

To further investigate differences in response time across the three groups (Research Question 3), standardized mean differences were calculated between accommodated and non-accommodated test items within student groups. Additionally, differences between student groups were examined on the accommodated items and on the non-accommodated items for each subject area. Standardized mean differences in response time were calculated using the formula below:

$$SMD = \frac{RT_1 - RT_2}{\sqrt{\frac{\sigma_1^2 + \sigma_1^2}{2}}} \tag{3.5}$$

where RT_1 and RT_2 are the natural logarithms of response time for each group, and σ_1^2 and σ_2^2 are the variances of response time for each group. Standardized mean differences of .20 or greater were considered of practical significance (Cohen, 1988).

3.5.2.4 Regression Analysis

To investigate which characteristics of ELs and non-ELs predict response time on accommodated and non-accommodated test items (Research Question 4), 6 multiple regression analyses for each subject area (History & Math) were completed. Specifically, separate analyses were conducted for each student subgroup (mid-proficient ELs, high-proficient ELs, & non-ELs) on Forms 1A and 2A (accommodated & non-accommodated test forms). Prior to running the analyses, a new variable was created in the data set indicating the average response time for each examinee on Forms 1A and 2A. The natural logarithm of each student group's average response time (Ln(RT)) was the dependent variable for each analysis. Predictors were the same predictors that were used to predict accommodation use and included: sex, ethnicity, socioeconomic status (SES), at-risk status, and statewide English Language Arts (ELA) score for non-ELs. Predictors for ELs included the same predictors as non-ELs and number of years in the United States and English language proficiency test score. It is important to note that the reason for running separate analyses for each EL group was because the EL group was likely to be highly correlated with number of years in the U.S. and English language proficiency test score thus causing multicollinearity. Running separate analyses removed this issue of multicollinearity. The formula for multiple regression is:

$$\underline{y} = X\underline{\beta} + \underline{e} \tag{3.6}$$

where $\underline{y}$ is the vector of average response times for each student, X is the score associated with each predictor (e.g., 0 for male, 1 for female), $\underline{\beta}$ is the vector of regression coefficients, and $\underline{e}$ is the vector of errors. Equation 3.6 is shown in matrix form below:

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1k} \\ 1 & x_{22} & \cdots & x_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n2} & x_{n3} & x_{nk} \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdots \\ e_n \end{bmatrix} \tag{3.7}$$

where n is the number of examinees and k is the number of predictors. In the case of non-ELs there were a total of 9 predictors. The sex, ethnicity, SES, and At-Risk variables were dummy coded (e.g., 1 = Black, 0 = All other ethnicities). Therefore, the regression equation for non-ELs on Form 1A was as follows:

$$\begin{aligned} \hat{y}_{AvgRT} &= \beta_0 + \beta_{Sex}X_{Sex} + \beta_{Black}X_{Black} + \beta_{Hisp}X_{Hisp} + \beta_{AmInd}X_{AmInd} \\ &+ \beta_{Asian}X_{Asian} + \beta_{NatHI}X_{NatHI} + \beta_{SES}X_{SES} + \beta_{AtRisk}X_{AtRisk} + \beta_{ELA}X_{ELA} \end{aligned} \tag{3.8}$$

These analyses helped to understand how specific subgroup characteristics can impact response times on accommodated and non-accommodated test items.

3.5.2.5 Item-Level Analyses

To examine the relationship between student proficiency, response time, and EL and non-EL accommodation status (Research Question 5), response time was investigated separately for each item on each subject test. Correlations were calculated between student raw score on the test and the mean Ln(RT) for each student group to see if there was a relationship between student performance on the 25-item assessment (original test Forms 1 & 2), and response time. Additionally, individual graphs were developed for each item. To do this, mean exp(Ln(RT)) was calculated at each raw score point for each student group, and plotted on a separate graph for each item with a different line for each accommodated and non-accommodated student group (6 groups total).

Using this information, student variation in response time patterns in relation to their performance on the assessment was examined. When interpreting the relationships, the number of accommodations on each item was considered to see how that impacted response time. Specifically, test items were re-ordered in relation to the number of accommodations on that particular item, and the mean exp(Ln(RT)) in seconds for each student group was graphed. Additionally, item difficulty for each item was examined to help explain why certain response time patterns on individual items could be occurring.

3.5.3 Summary of Analyses

A summary table with the research questions and corresponding analyses is shown in Table 3.3. This table indicates the analysis level (test or item-level), the test forms involved in the analysis, which analysis was conducted, and the variables used in the analysis. It is important to note that all analyses shown in Table 3.3 were conducted for both subject areas.

Table 3.1. History Demographic Information by Form

	Form 1			Form 2		
	Mid-Prof ELs	High-Prof ELs	Non-ELs	Mid-Prof ELs	High-Prof ELs	Non-ELs
Number of Students	297	447	547	284	485	505
Males	50.2%	50.3%	43.7%	53.9%	51.3%	46.1%
Hispanic	77.8%	70.9%	48.1%	78.2%	72.2%	49.7%
American Indian/Alaskan Native	1.0%	.4%	.4%	1.1%	1.2%	.2%
Asian	.7%	1.3%	4.0%	1.4%	0.4%	2.6%
Black/African American	14.8%	17.9%	3.8%	11.3%	16.7%	3.4%
Native Hawaiian/Pacific Islander	2.4%	5.4%	.2%	4.9%	4.7%	.2%
White	.3%	0%	26.5%	.4%	.2%	27.1%
Free/Reduced Price Lunch/Low SES	87.8%	85.8%	49.9%	88.7%	86.3%	46.6%
Bilingual Program	.7%	0.0%	0%	0%	.6%	0%
ESL Program	97.0%	98.0%	0%	97.9%	96.9%	0%
SPED Program	.7%	1.3%	2.6%	0%	1.2%	3.2%
At-Risk	98.3%	97.8%	41.5%	95.8%	97.9%	41.8%
Passed ELA State Test	1.0%	18.1%%	100%	1.8%	19.0%	100%

Table 3.2. Math Demographic Information by Form

	Form 1			Form 2		
	Mid-Prof ELs	High-Prof ELs	Non-ELs	Mid-Prof ELs	High-Prof ELs	Non-ELs
Number of Students	292	475	368	276	453	328
Males	52.1%	50.5%	50.5%	52.5%	53.2%	47.3%
Hispanic	76.0%	70.1%	20.7%	76.8%	72.8%	17.1%
American Indian/Alaskan Native	1.0%	.6%	.3%	1.1%	.9%	.3%
Asian	0%	1.7%	1.4%	1.1%	.2%	1.8%
Black/African American	16.4%	17.5%	24.2%	13.4%	17.0%	20.7%
Native Hawaiian/Pacific Islander	3.1%	5.2%	0%	3.3%	3.8%	0%
White	.3%	0%	45.4%	0%	.4%	54.9%
Free/Reduced Price Lunch/Low SES	87.4%	85.5%	29.4%	87.7%	85.4%	27.4%
Bilingual Program	.7%	.6%	0%	.4%	0%	0%
ESL Program	98.7%	98.5%	0%	97.9%	96.9%	0%
SPED Program	1.0%	.6%	4.9%	.4%	.7%	3.7%
At-Risk	99.3%	97.9%	31.5%	96.7%	96.5%	28.7%
Passed ELA State Test	2.4%	16.4%	100%	1.1%	15.9%	100%

Table 3.3. Analysis Summary Table

RQ	Category	Analysis Level	Analysis	Variables
1	Accom Use	Test Level	One-Way ANOVA	Total accom use by student group
		Item-Level	Chi-Square	Accom use by student group
			Plot	Accom use by item location
				Frequencies of accom use by student group
2	Accom Use	Test Level Form 1A	Poisson Regression	Predictors: Sex, ethnicity, SES, at-risk status, ELA score, # years in U.S. (EL only), English proficiency test score (EL only)
3	Response Time	Test Level Forms 1A, 2A	Two-Way ANOVA	DV: Mean Ln(RT) per item IVs: Test form, student group
		Item-Level Forms 1A, 2A	SMD	Response time differences by student group
4	Response Time	Test Level Forms 1A, 2A	Multiple Regression	Predictors: Sex, ethnicity, SES, at-risk status, ELA score, # years in U.S. (EL only), English proficiency test score (EL only)
5	Response Time	Item-Level	Correlation	Raw score and Mean Ln(RT)
			Plot	Mean Ln(RT) by raw score for each student group

Note. RQ = Research question; Accom = Accommodation; SMD = Standardized mean difference.

Item #	Form 1	Form 2
1	*Accommodation*	No Accommodation
2	No Accommodation	*Accommodation*
3	*Accommodation*	No Accommodation
…	…	…
25	No Accommodation	*Accommodation*

Figure 3.1. Accommodated vs. Not Accommodated Test Items Across Forms

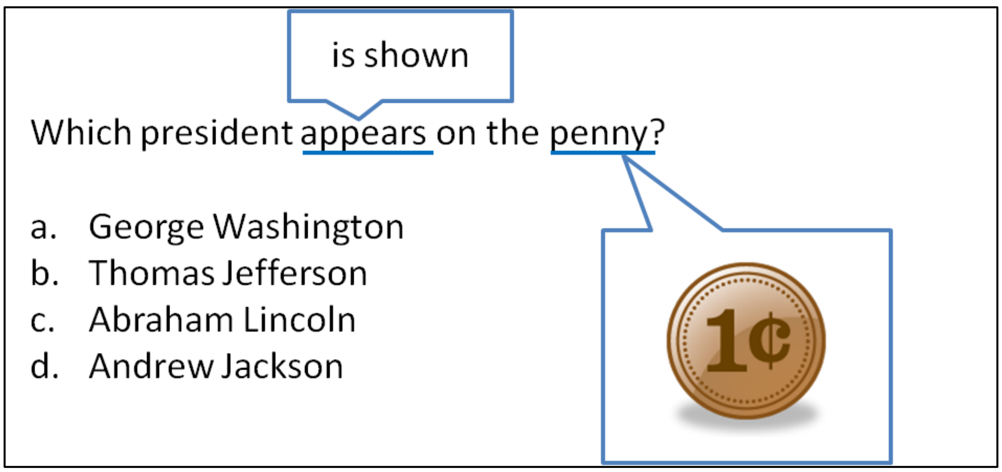


Figure 3.2. Example of Pop-Up Glossary Tool

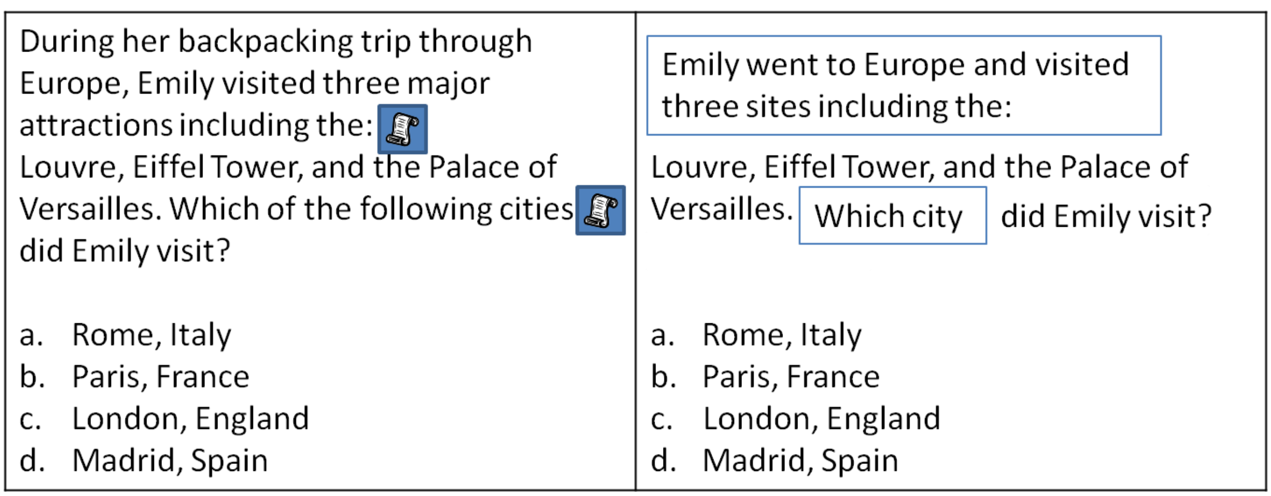


Figure 3.3. Example of Sticker Paraphrasing Tool

Item #	Form 1A	Form 2A
1	*Accommodation*	No Accommodation
2	*Accommodation*	No Accommodation
3	*Accommodation*	No Accommodation
…	…	…
23	*Accommodation*	No Accommodation
24		No Accommodation
25		No Accommodation

Figure 3.4. Example of the New Test Forms Used for Analyses

	Accommodation Use			
	0	1	2	3
Mid-proficient ELs				
High-Proficient ELs				
Non-ELs				

Figure 3.5. Example of Contingency Table for Individual Test Items

CHAPTER 4

RESULTS

4.1 Overview

This section presents the results of the methodology discussed in Chapter 3. Each of the five research questions were addressed for the two subject areas. Results are presented for History followed by Mathematics. Performance descriptive statistics are provided first, followed by the accommodation use analysis results, and the response time analysis results. Accommodation use results include differences in accommodation use across student groups (Research Question 1), and characteristics predicting accommodation use (Research Question 2). Response time results include descriptive statistics, differences across student groups (Research Question 3), characteristics predicting response time (Research Question 4), and the relationship between student proficiency, response time, and accommodation use (Research Question 5). Additionally, the relationship between accommodation use and response time is examined, finishing with a summary of the statistically significant results for each subject area.

4.2 History

4.2.1 Performance Descriptive Statistics

Table 4.1 shows the History performance descriptive statistics across Forms 1 and 2 for the three student groups, removing any students identified as receiving special education. Forms 1 and 2 were the original forms administered to students with accommodations alternating every other item. Results indicated that non-EL students performed highest on both test Forms, followed by high-proficient ELs, and mid-proficient ELs. Performance across both test forms was similar for the three student groups, which was expected due to the random assignment of test forms.

4.2.2 Differences in Accommodation Use Between Student Groups

To investigate overall differences in accommodation use between the non-EL and EL student groups, a one-way ANOVA was conducted with total accommodation use as the dependent variable. Total accommodation use was relative to which test form students took. In total, there were 53 accommodations available on Form 1 and 60 accommodations available on Form 2, resulting in a total of 113 accommodations among the 23 accommodated test items. Results yielded statistically significant differences in accommodation use for the three student groups ($F_{(2, 2518)} = 142.31$, $p < .001$, $\eta^2 = .102$). Specifically, statistically significant differences were found between non-ELs and high-proficient ELs ($t_{(2518)} = 14.13$, $p < .001$, $d = .662$), and between non-ELs and mid-proficient ELs ($t_{(2518)} = 14.24$, $p < .001$, $d = .756$). Differences in accommodation use between the two EL groups were not significant.

To further investigate the differences in accommodation use between the non-EL and EL student groups, chi-square analyses were completed separately for each test item taking into account the different number of accommodations available on each item. Table 4.2 shows the results, indicating that across the 23 History items with test accommodations, all items showed statistically significant relationships between student group and the number of accommodations used on that item ($p < .001$). Cramer's V results indicated that the strength of the relationship between student group and accommodation use ranged from .124 and .264 indicating small to somewhat moderate relationships between the two variables. As mentioned in Chapter 3, Cramer's V varies between 0 indicating no association, and 1 indicating a strong association. Moderate associations begin at .20 (Rea & Parker, 1992). Approximately 11 of the 23 items had a Cramer's V of .20 or higher.

When investigating why certain items showed larger associations between accommodation use and student group, patterns across the test items were explored. The most notable pattern in the size of Cramer's V was found in relation to item location. Specifically, items with a stronger relationship between student group and accommodation use tended to be towards the end of the assessment. Items 19-25, except item 21, all had a Cramer's V of .20 or higher. The remaining five items with a Cramer's V of .20 or higher all had 6 accommodations or more, which likely impacted the differences in accommodation use. Lastly, to investigate whether item difficulty was related to accommodation use, proportion correct on the items for each student group was examined. It was noted that items showing larger differences in accommodation use were the same items that showed large differences in proportion correct ($\geq .20$) between non-ELs and the two EL student groups.

In addition to the chi-square analyses, graphs were created showing the frequencies of accommodations for each individual item (see Appendix A). When examining these graphs it was clear that non-ELs used accommodations less frequently than the EL groups. Specifically, at least 60% of non-ELs either used zero or one accommodation on most items, with the majority of non-ELs not using any of the accommodations. Four items (Items 5, 9, 11, 16) showed at least 40% of non-ELs either using zero or one accommodation, which could have been due to the high number of accommodations on those specific items (6, 9, 11, & 5 accommodations, respectively).

On almost all of the test items, 60% of mid-proficient ELs used at least one accommodation, with the majority of mid-proficient ELs using all available accommodations. High-proficient ELs showed similar trends to mid-proficient ELs, which was expected since differences in accommodation use between those two groups were not statistically significant.

To examine whether accommodation use was linked to item location, a graph was created showing the percentage of students not using an accommodation on a specific item by the sequential location of the items on the assessment. As shown in Figure 4.1, there was a slight increase in percentage of students not accessing accommodations for all student groups towards the end of the assessment. Of the three student groups, non-ELs were least likely to use an accommodation on an individual item. Specifically, on an individual item, 17-19% of non-ELs did not use any accommodations as compared to ELs. The number of ELs not using accommodations was similar across both high-proficient and mid-proficient ELs.

4.2.3 Characteristics Predicting Accommodation Use

Before completing the Poisson regression, the assumption in relation to the conditional mean and variance of the accommodation use variable was checked. Because separate regressions for each student group were being completed, the accommodation use variable was examined separately for each group to see if the assumption was met. When completing the Poisson regressions, it was noted that mean and variances were not equal; therefore, negative binomial regressions should be used. Negative binomial regressions do not make the assumption that the mean and variance are equal, and instead correct for the overdispersion of the data (Piza, 2012). Similarly, the model fit using the

Poisson regression was very poor, also indicating that the negative binomial regression be used instead.

Table 4.3 shows the regression coefficients for the three negative binomial regressions used to predict accommodation use for the three different student groups. Using the negative binomial regression, all three regression models fit the data with a deviance ratio around 1.00. The non-EL model yielded statistically significant predictors for sex, Hispanic, American Indian, and ELA score. To interpret these predictors, the incident rate ratio was looked at by taking the base *e* of the coefficient, *B*. With respect to sex, males used accommodations 20% less than females. Among the ethnic predictors, Hispanic students were 34% more likely to use the accommodations, and American Indian students were 87% less likely to use the accommodations.

Across the two EL groups, results indicated that for the high-proficient EL model, statistically significant predictors included sex, Hispanic, and SES. Results indicated that males were 22% less likely to use accommodations than females, Hispanic students were 41% less likely to use accommodations, and students with low-SES were 50% more likely to use accommodations. For the mid-proficient EL model, the model itself was not statistically significant (p = .290), and as a result there were no statistically significant predictors of accommodation use for that student group.

4.2.4 Response Time Descriptive Statistics

Table 4.4 shows the descriptive statistics for response time across the two original test forms (Forms 1 & 2). Specifically, median response time in minutes, and mean and standard deviation of the natural logarithm of total response time (Ln(RT)) were examined because response time is positively skewed. To interpret Ln(RT), response time was converted back to minutes by taking the base of the natural logarithm, exp(Ln(RT)). Results indicated that each student group completed the assessment using approximately the same amount of time regardless of test form. Across student groups, non-ELs took the least amount of time to complete the assessment, followed by high-proficient ELs and mid-proficient ELs, which was expected. Additionally, the spread in response time was quite large for both EL groups with standard deviations (SDs) between 18 and 22.

Because of interest in differences in response time across accommodated and non-accommodated test items, the response time across Forms 1A (accommodated) and 2A (non-accommodated) was also examined. Because these forms were not the intact forms taken by examinees, but instead were created for the purpose of this comparison, the median of median response times (in seconds) on each item across Forms 1A and 2A was examined, as well as the mean and SD of exp(Ln(RT)) for the three groups. Table 4.5 shows that accommodated items (Form 1A) took longer to complete for all student groups as compared to non-accommodated items. Similar to the total response times shown in Table 4.4, non-ELs took the least amount of time on an individual item, followed by high-proficient ELs and mid-proficient ELs.

Appendix B shows response times for each individual item on Forms 1A (Table B.1) and 2A (Table B.2). Specifically, both tables show the median response time in seconds, and the mean and SD of exp(Ln(RT)) in seconds. Differences in amount of time across accommodated and non-accommodated test items on average was around 5 seconds for non-ELs, 10 seconds for high-proficient ELs, and 12 seconds for mid-proficient ELs, meaning students took slightly longer on accommodated items as compared to the same non-accommodated items. High-proficient ELs took about 22 and 19 seconds longer on accommodated and non-accommodated items, respectively, as compared to non-ELs. Similarly, mid-proficient ELs took about 28 seconds longer on average on accommodated items, and 24 seconds longer on non-accommodated items when compared to non-ELs. Additionally, mid-proficient ELs took about 6 and 4 seconds longer on accommodated and non-accommodated items, respectively, as compared to high-proficient ELs.

4.2.5 Differences in Response Time across Student Groups

A two-way ANOVA was conducted to examine whether ELs and non-ELs take statistically significantly longer to complete items than ELs and non-ELs without an accommodation. Table 4.6 shows the ANOVA summary table. Results indicated a statistically significant main effect for student group ($p < .001$, $\eta^2 = .441$) with a very large effect size. Specifically, statistically significant differences were found between non-ELs and both the mid-proficient ($t_{(138)} = -9.75$, $p < .001$, $d = -1.660$), and high-proficient ELs ($t_{(138)} = -8.56$, $p < .001$, $d = -1.457$); however, there was no statistically significant difference in response time between the two EL groups ($t_{(138)} = 1.21$, $p = .230$). When controlling for family wise error using Fisher LSD, the same statistically significant differences were found.

A statistically significant main effect was also found for test form (Forms 1A & 2A) ($p = .014$, $\eta^2 = .024$) with a small effect size, indicating a statistically significant but small difference in response time across the accommodated and non-accommodated test items. There was no statistically significant interaction meaning that the same patterns in response time were found for the student groups across the accommodated and non-accommodated test forms.

In addition to the two-way ANOVA, the standardized mean response time difference was examined between accommodated and non-accommodated test items, and between the student groups. Table 4.7 shows the item-level standardized mean differences. Standardized mean differences of .20 or greater are considered of practical significance (Cohen, 1988). Results between the accommodated and non-accommodated items showed many of the items taking longer with accommodations than without accommodations. Specifically, 9, 10, and 6 items showed practically significant differences in response time for non-ELs, mid-proficient ELs, and high-proficient ELs, respectively (see Table 4.8). In total, there were 14 unique items that showed practical significance. Of these 14 items, 4 were common between all three student groups. One additional item was common between non-ELs and high-proficient ELs, and two additional items were common between non-ELs and mid-proficient ELs. Typically, items with more accommodations showed larger differences in response time, and items with fewer numbers of accommodations showed smaller differences in response time. Similarly, items without accommodations across both test forms all showed differences close to zero.

Results indicated large standardized mean differences of practical significance between non-ELs and the two EL groups on both the accommodated and non-accommodated test items, with EL students taking statistically significantly longer on individual items as compared to non-ELs. Differences in response time between the two EL groups were very small, with only 2 items showing practical significance on the accommodated items, and one item showing practical significance on the non-accommodated items. The three items showing practical significance showed mid-proficient ELs taking significantly longer to respond to items than high-proficient ELs.

4.2.6 Characteristics Predicting Response Time

A summary of the regression results is presented in Table 4.9. Six separate regressions were completed for three different student groups across test Forms 1A and 2A. On the accommodated form (Form 1A), the most variance in response time was explained for the high-proficient ELs (20%), followed by mid-proficient ELs (14%), and non-ELs (11%). Sex was a statistically significant predictor across all three student groups with males taking less time to respond to items than females. Black status was statistically significant in both the non-EL and mid-proficient EL regressions with Black students taking longer to respond to items. Hispanic and Asian predictors were also statistically significant and positive for the non-EL regression, and Native Hawaiian was statistically significant and positive for the mid-proficient EL regression. Interestingly, for the high-proficient ELs, both the Hispanic and American Indian ethnic predictors were statistically significant but negative, indicating that those respective groups took less time to respond to items, the opposite of the results found for the non-EL and mid-proficient EL regressions. Additionally, SES, at-risk status, and ELA score were all statistically significant for the high-proficient EL regression on Form 1A (accommodated form).

Results for the non-accommodated test form (Form 2A) indicated that less variance was explained for both EL groups at 17% and 10% for the high-proficient and mid-proficient ELs, respectively. For non-ELs, 11% of the variance in response time was explained with the same statistically significant predictors that were found in the Form 1A analysis. For high-proficient ELs, sex, Hispanic, SES, ELA score, and ELP score were all statistically significant. Results indicated lower response time rates for males, Hispanic students, and ELs with higher ELP scores. The only statistically significant predictor for mid-proficient ELs on Form 2A was Black.

4.2.7 Relationship between Student Proficiency and Response Time

4.2.7.1 Correlation

To examine the relationship between student proficiency and response time, Pearson correlations were computed between the test raw score and the average Ln(RT) for Forms 1 and 2, and for the three student groups. Results indicated that on both test forms non-ELs had a statistically significant moderately positive relationship between raw test score and average response time (see Table 4.10). For the EL groups, however, relationships were positive and quite small. Additionally, the relationships for the EL groups were only statistically significant on Form 1, and not Form 2.

To further analyze these correlations, correlations across the student groups on the two test forms were tested for statistically significant differences. To do this, the Fisher *z'* transformation of *r* was completed using the formula below:

$$z' = \frac{1}{2}\left[\ln(1+r) - \ln(1-r)\right] \tag{4.1}$$

Once the transformation was completed, statistical differences between *z*-scores were tested by computing the normal curve deviate (Cohen et al., 2003):

$$z = \frac{z'_1 - z'_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}, \tag{4.2}$$

where *n* is the sample size for each student group. Results indicated statistically significant differences in correlations between non-ELs and both EL groups with small effect sizes on both test Forms 1 and 2 (see Table 4.10). Correlational differences between high-proficient and mid-proficient ELs on both test forms were not statistically significant.

4.2.7.2 Item-Level Analysis

Appendix C shows the graphs of raw score by the exp(Ln(RT)) reported in seconds. Average response time was calculated for students in each student group at each raw score point. If a raw score point contained less than 10 examinees, it was not plotted in the graph. On all items, there was a clear trend of non-ELs taking the least amount of time, and the EL groups taking more time to respond to items. Additionally, on many of the items, students receiving an accommodation took slightly longer to complete the item than students without an accommodation, which was consistent with the ANOVA results indicating a statistically main effect for test form (p = .014, η^2 = .024). Across all test items, non-ELs showed either stable response time across the raw score scale, or an increase in response time with higher performance. In total, 14 items showed non-EL response time increasing with increased performance on the test. For both EL groups, response time was much more jagged across the score scale than non-EL response time. The majority of items showed jagged, but stable response time across the score scale; however, about 8 items showed some increase in response time with increased performance.

4.2.8 Relationship between Response Time and Accommodation Use

To investigate the relationship between response time and accommodation use, test items were re-ordered in relation to the number of accommodations on that particular item (ranging from 1 to 11 accommodations), and the mean exp(Ln(RT)) in seconds for each student group was graphed. As shown in Figure 4.2, average student response time increased for both EL groups as number of accommodations increased, but remained fairly stable for non-ELs.

4.2.9 Summary of History Results

Table 4.11 gives a summary of the statistically significant results for accommodation use. Statistically significant differences in accommodation use were found across all accommodated test items. Additionally, 11 accommodated items showed moderate associations in relation to differences in accommodation use. These 11 accommodated items tended to be towards the end of the assessment, and showed large differences in item difficulty across student groups. In relation to predictors of accommodation use, statistically significant predictors for non-ELs included sex, Hispanic status, American Indian Status, and ELA score. Statistically significant predictors for high-proficient ELs included sex, Hispanic status, and low-SES. There were no statistically significant predictors of accommodation use for mid-proficient ELs.

Table 4.12 gives a summary of the statistically significant results for response time. Statistically significant differences in response time were found across student groups, specifically between non-ELs and the EL groups, and were found across accommodated and non-accommodated test forms. Looking at the standardized mean differences in response time across accommodated and non-accommodated items for the student groups indicated that 9, 6, and 10 items showed practically significant differences in response time for non-ELs, high-proficient ELs, and mid-proficient ELs, respectively. Practically significant differences were found on all items between non-ELs and both EL groups on the accommodated and non-accommodated test items. Only 2 accommodated items and 1 non-accommodated item showed practically significant differences in response time between high-proficient and mid-proficient ELs.

In relation to predictors of response time, statistically significant predictors for non-ELs included sex, Black status, Hispanic status, and Asian status on both accommodated and non-accommodated test forms (Forms 1A & 2A). For high-proficient ELs, significant predictors on the accommodated test form (Form 1A) included sex, Hispanic status, American Indian status, SES, at-risk, and ELA score. On the non-accommodated test form (Form 2A), statistically significant predictors included sex, Hispanic, SES, ELA score, and ELP score. For mid-proficient ELs, statistically significant predictors included Black status, Native Hawaiian status, and sex for the accommodated test form (Form 1A), and only Black status for the non-accommodated test form (Form 2A).

Table 4.12 also shows the significant results for the relationship between response time and raw score. Statistically significant correlations were found on Form 1 for all student groups, and only for non-ELs on Form 2. Statistically significant differences in the correlations were found between non-ELs and both EL student groups on both test forms.

4.3 Mathematics

4.3.1 Performance Descriptive Statistics

Table 4.13 shows the Mathematics performance descriptive statistics across Forms 1 and 2 for the three student groups, again removing any students identified as receiving special education. Results indicated that non-EL students performed highest on both test Forms, followed by the high-proficient ELs, and mid-proficient ELs. Performance across both test forms was similar for the three student groups, which was expected due to the random assignment of test forms. Forms 1 and 2 were the original forms administered to students with accommodations alternating every other item.

4.3.2 Differences in Accommodation Use Between Student Groups

A one-way ANOVA was conducted to identify which groups showed statistically significant differences in accommodation use throughout the assessment. For Math, there were a total of 45 accommodations available on Form 1, and 42 accommodations available on Form 2, resulting in a total of 87 accommodations available across the 19 accommodated test items. Results yielded statistically significant differences in accommodation use across the assessment ($F_{(2, 2149)} = 100.46$, $p < .001$, $\eta^2 = .085$). The statistically significant differences were between non-ELs and high-proficient ELs ($t_{(2149)} = 10.71$, $p < .001$, $d = .596$), and between non-ELs and mid-proficient ELs ($t_{(2149)} =$ 13.48, $p < .001$, $d = .798$). Statistically significant differences in accommodation use were also found between both EL groups ELs ($t_{(2149)} = 4{,}24$, $p < .001$, $d = .204$).

To further investigate differences in accommodation use across the three student groups, item-level chi-square analysis were conducted. Results for the item-level accommodation use differences between non-EL and EL student groups across the 19 accommodated math items indicated statistically significant relationships between student group and accommodation use on 18 of the 19 items ($p < .001$) (see Table 4.14). Cramer's V results indicated that the strength of the relationship between student group and accommodation use on the statistically significant items ranged from .129 to .309 indicating small to moderate relationships between the two variables. A total of 7 of the 19 items had a Cramer's V of .20 or higher.

To examine if there were any potential patterns in the size of Cramer's V, both the number of accommodations and the differences in proportion correct on the items for each student group were investigated. Similar to the History results, items with a stronger relationship between student group and accommodation use tended to be towards the end of the assessment (items 18, 20, 24, 25). Of the items that were not towards the end of the assessment, all of the items had at least 5 accommodations available, suggesting that number of accommodations available on an item did play a role unless the item was towards the end of the assessment. In relation to proportion correct (item difficulty), it was noted that the two items with the highest Cramer's V were difficult items for all student groups.

Appendix D shows the frequencies of accommodation use for each individual item. Results for non-ELs indicated that accommodation use was rare, with at least 50% of non-ELs not using accommodations on any of the items. On almost all of the test items, at least half of the mid-proficient ELs used at least one accommodation, with the majority using all available accommodations. High-proficient ELs showed similar trends to mid-proficient ELs, however, the percentage of students using accommodations on each item was slightly lower in relation to mid-proficient ELs.

To investigate if there were any trends in relation to accommodation use and item location, a graph was created showing the item number by the frequency of zero accommodation use on a single item. As shown in Figure 4.3, the percentage of students not using accommodations did have a slight increase throughout the assessment. Results indicated that non-ELs were least likely to use accommodations, followed by high-proficient ELs, and mid-proficient ELs. Specifically, on all of the items, 17-22% of non-ELs did not use any accommodations as compared to the two EL groups.

4.3.3 Characteristics Predicting Accommodation Use

Similar to the History assessment, before completing the Poisson regression, the assumption in relation to the conditional mean and variance of the accommodation use was examined separately for each group, and it was noted that the assumption was not met. Similarly, model fit using the Poisson regression was very poor indicating that a negative binomial regression be used instead to determine predictors of accommodation use for the Math assessment. Table 4.15 shows the regression coefficients for the three negative binomial regressions used to predict accommodation use for the three different student groups. Results indicated that all three regressions had moderate fit to the data with deviance ratios close to 1.50.

Results of the non-EL model indicated sex was a statistically significant predictor, with males 20% more likely to use accommodations as compared to females. The only statistically significant predictor of accommodation use for the high-proficient EL model was SES. Students with low-SES were 26% more likely to use accommodations than students not identified as low-SES. For the mid-proficient EL model, statistically significant predictors included Black, Hispanic, Native Hawaiian, and at-Risk student status. Specifically, Black, Hispanic, and Native Hawaiian students were 130%, 89%, and 184% more likely to use the accommodations, respectively. Additionally, at-risk students were 323% more likely to use accommodations than students who were not considered at-risk.

4.3.4 Response Time Descriptive Statistics

Table 4.16 shows the descriptive statistics for response time across the two original test forms (Forms 1 & 2). Across both test forms non-ELs took the least amount of time, completing the assessment in approximately 30 minutes. Both EL groups took quite a bit longer, taking around 55 minutes. Additionally, the spread in response time was quite large for all three groups with SDs around 15 for non-ELs, and 29 for the two EL groups. Response time was consistent across test forms. Across accommodated and non-accommodated test items (Forms 1A & 2A), results showed longer response times for accommodated items as compared to non-accommodated items (see Table 4.17). Similar to total test time, non-ELs took the least amount of time on individual items, followed by high-proficient ELs, and mid-proficient ELs.

Appendix E shows the response times for each individual item on Forms 1A (Table E.1) and 2A (Table E.2). Differences in amount of time across accommodated and non-accommodated test items were around 9 seconds for non-ELs and 14 seconds for both high- and mid-proficient ELs. Across student groups, non-ELs took 46 seconds and 49 seconds less than high- and mid-proficient EL groups, respectively on accommodated items, and 38 and 40 seconds less than high- and mid-proficient EL groups, respectively, on non-accommodated items. High-proficient ELs took 3 seconds less than mid-proficient ELs on accommodated items, and 2 seconds less on non-accommodated items.

4.3.5 Differences in Response Time across Student Groups

Results of the two-way ANOVA examining whether ELs and non-ELs take statistically significantly longer to complete items with an accommodation than ELs and non-ELs without an accommodation can be found in Table 4.18. A statistically significant main effect was found for student group with a large effect size ($p < .001$, $\eta^2 = .201$). Specifically, statistically significant differences were found between non-ELs and both the high-proficient ELs ($t_{(126)} = -4.76$, $p < .001$, $d = -.848$) and mid-proficient ($t_{(126)} = -5.15$, $p < .001$, $d = -.918$); however, there was no statistically significant difference in response time between the two EL groups ($t_{(126)} = .396$, $p = .690$). The same statistically significant differences were found when controlling for family wiser error using Fisher LSD.

A statistically significant main effect was also found for test form (Forms 1A & 2A) ($p = .027$, $\eta^2 = .031$) with a small effect size, indicating a statistically significant but small difference in response time across the accommodated and non-accommodated test items, with the accommodated items taking longer. There was no statistically significant interaction meaning that the same patterns in response time were found for the student groups across the accommodated and non-accommodated test forms.

Table 4.19 shows the item-level standardized mean differences between accommodated and non-accommodated test items, and between the student groups. Results indicated very small differences in response time between accommodated and non-accommodated items across the three groups with only one item showing practical significance for non-ELs and mid-proficient ELs, and 3 items showing practical significance for high-proficient ELs (see Table 4.20). In total, there were 3 unique items that showed practical significance, with the 1 item being common between all three student groups. The two additional items were only unique for the high-proficient ELs. Among the three items showing practical significance, differences indicated that the accommodated items took slightly longer for students. Small to moderate practically significant differences were found between non-ELs and high-proficient ELs on accommodated items, and moderate to large practically significant differences were found between non-ELs and mid-proficient ELs. Differences between the EL groups were very small and therefore non-significant. Only one non-accommodated item showed a practically significant difference with mid-proficient ELs taking longer than high-proficient ELs.

4.3.6 Characteristics Predicting Response Time

Six separate regressions were completed for three student groups across test forms 1A and 2A. Results for the accommodated test Form 1A showed that the most variance explained in response time was for the high-proficient ELs at 13%, followed by mid-proficient ELs (10%), and non-ELs (3%) (see Table 4.21). Across the Form 1A (accommodated form) models, sex was a statistically significant predictor for both the non-EL and mid-proficient EL models, with males taking less time to respond to items than females. Black status was a statistically significant predictor of response time for all three groups, with Black students taking longer on items than other ethnic groups. Additionally, ELP score was a statistically significant predictor for high-proficient ELs, and Native Hawaiian was a statistically significant predictor for mid-proficient ELs. Response time was higher for Native Hawaiian students, and decreased for students as ELP score increased.

Results for Form 2A (non-accommodated form) were very similar to the Form 1A (accommodated form) results, explaining similar amounts of variance in response time at 12%, 11%, and 4%, for mid-proficient ELs, high-proficient ELs, and non-ELs, respectively. Again, sex and Black status were statistically significant predictors for both the non-EL and mid-proficient EL models. Additionally, both Asian and Native Hawaiian status were statistically significant predictors for the mid-proficient EL model, with Asian students taking less time to respond to items compared to the other ethnic groups. The only two statistically significant predictors for the high-proficient EL model were Hispanic and ELP score, with Hispanic students taking less time to respond to items.

4.3.7 Relationship between Student Proficiency and Response Time

4.3.7.1 Correlations

Table 4.22 shows the relationship between student proficiency and response time. Results indicated that on both test forms non-ELs had a statistically significant moderately positive relationship between raw score and average Ln(RT). For high-proficient ELs, the relationship between raw score and average response time was small, but statistically significant. For mid-proficient ELs, results showed a very small statistically significant relationship on Form 1, but not on Form 2.

Similar to the History results, to further analyze the correlations, differences between correlations were tested across the student groups on the two test forms first using the Fisher *z'* transformation of *r* (see Formula 4.1), then testing for statistical differences between *z*-scores by using the normal curve deviate (see Formula 4.2). Results indicated statistically significant differences in correlations between non-ELs and both EL groups on Form 1 with a small to moderate effect size (see Table 4.22).

Differences between the two EL groups were not statistically significantly different on Form 1. On Form 2, however, statistically significant differences in correlations were found between non-ELs and mid-proficient ELs, and between high-proficient and mid-proficient ELs. There were no statistically significant differences in correlations between non-ELs and high-proficient ELs on Form 2.

4.3.7.2 Item-Level Analysis

Appendix F presents the graphs of raw score by the exp(Ln(RT)) reported in seconds. Similar to the History assessment, if a raw score point contained less than 10 examinees, it was not plotted in the graph. On almost all items there was a clear trend of non-ELs taking the least amount of time, and EL groups taking more time to respond to items. Exceptions included items 8 and 14, which were both non-accommodated items. There was a trend on some of the items where students receiving an accommodation took longer to complete an item than students without an accommodation, but this trend was inconsistent. About 15 of the items showed a trend of response time increasing as raw score increased. Items towards the end of the assessment (Items 20 through 25) had the most stable response times across raw score.

4.3.8 Relationship between Response Time and Accommodation Use

Figure 4.4 shows the relationship between response time and accommodation use. Items were reordered in relation to the number of accommodations on a particular item (this ranged from 1 to 8 accommodations). No consistent trend was found in relation to response time and the increased number of accommodations on an individual item.

4.3.9 Summary of Mathematics Results

Table 4.23 gives a summary of the statistically significant results for accommodation use. Statistically significant differences in accommodation use were found across 18 of the 19 accommodated items. In total, 7 accommodated items showed moderate associations in relation to accommodation use and student group. In relation to statistically significant predictors of accommodation use, results for non-ELs indicated that only sex was a statistically significant predictor. For high-proficient ELs, only SES was a statistically significant predictor of accommodation use. For mid-proficient ELs, statistically significant predictors included at-risk, Black status, Native Hawaiian status, and Hispanic status.

Table 4.24 gives the summary of statistically significant results in relation to response time analyses. Statistically significant differences in response time were found across student groups, specifically between non-ELs and the EL groups, and were found across accommodated and non-accommodated test forms. Looking at the standardized mean differences in response time across accommodated and non-accommodated items for the student groups indicated that only 1, 3, and 1 items showed practically significant differences in response time for non-ELs, high-proficient ELs, and mid-proficient ELs, respectively. Practically significant differences were found on all, except 2 of the items between non-ELs and both EL groups on the accommodated test items. Similarly, 2 and 1 items showed practically significant differences in response time between non-ELs and high-proficient and mid-proficient ELs, respectively on the non-accommodated test items. Only one non-accommodated item showed practically significant differences in response time between high-proficient and mid-proficient ELs.

Statistically significant predictors of response time for non-ELs on both accommodated and non-accommodated items included sex and Black status. Both of these predictors were also statistically significant for mid-proficient ELs. Other statistically significant predictors for mid-proficient ELs included Native Hawaiian status on both accommodated and non-accommodated test forms, and Asian status on the non-accommodated test form. For high-proficient ELs, Black status and ELP score were statistically significant predictors for accommodated items, and Hispanic status and ELP score were statistically significant predictors for non-accommodated items.

Table 4.24 also shows the significant results for the relationship between response time and raw score. Statistically significant correlations were found on Form 1 for all student groups, and for non-ELs and high-proficient ELs on Form 2. Statistically significant differences in the correlations were found between non-ELs and high-proficient ELs on Form 1, and between non-ELs and mid-proficient ELs on both test forms. Statistically significant differences in correlations were also found between high-proficient and mid-proficient ELs on Form 2.

Table 4.1. History Test Score Descriptive Statistics

Form	Student Group	N	Mean	SD
1	Non-EL	533	14.17	4.59
	High-Prof EL	441	10.27	2.92
	Mid-Prof EL	295	9.02	2.66
2	Non-EL	489	14.11	4.50
	High-Prof EL	479	10.38	2.98
	Mid-Prof EL	284	9.45	2.81
	All Students	2521	11.63	4.22

Note. Prof = Proficient; SD = Standard deviation

Table 4.2. History Differences in Accommodation Use Between Student Groups

Item[a]	# Accoms	χ^2	df	Cramer's V
20	2	176.29[b]	4	.264
18	7	175.02	14	.263
24	2	175.71	4	.263
23	2	161.45	4	.254
25	4	157.50	8	.251
22	6	137.33	12	.233
6	9	136.04	18	.232
2	8	126.54	16	.223
19	6	121.50	12	.220
12	6	113.70	12	.212
11	11	110.38	22	.210
10	1	49.12	2	.197
15	8	91.95	16	.192
4	4	92.91	8	.191
9	9	90.66	18	.190
21	5	85.77	10	.185
17	3	82.29	6	.181
16	5	82.40	10	.180
5	6	66.43	12	.163
8	3	51.29	6	.142
13	2	48.98	4	.140
7	3	45.33	6	.135
1	1	19.24	2	.124

Note. Accoms = # of accommodations available for the item; df = Degrees of freedom.
[a]Items 3 and 14 did not have test accommodations and are not included in the table
[b]All χ^2 were significant ($p < .001$)

Table 4.3. History Negative Binomial Regression Coefficients Predicting Accommodation Use

	Non-EL			High-Proficient EL			Mid-Proficient EL		
	B	Exp(B)	SE	B	Exp(B)	SE	B	Exp(B)	SE
Intercept	4.92*	137.09	.89	5.27*	194.65	.92	3.54*	34.45	.89
Sex	-.22*	.80	.08	-.25*	.78	.08	-.17	.85	.09
Hispanic	.29*	1.34	.08	-.53*	.59	.20	.03	1.03	.31
Am Indian	-2.07*	.13	.77	-.76	.47	.55	.07	1.08	.78
ELA Score	-.001*	1.00	.00	.00	1.00	.00	.00	1.00	.00
SES	-.08	.92	.08	.41*	1.50	.11	.12	1.13	.15
Black	.11	1.11	.19	-.03	.97	.21	.37	1.45	.33
Asian	.13	1.14	.19	-.38	.68	.43	-.29	.75	.67
Native HI	.42	1.52	.72	-.13	.87	.26	.36	1.43	.39
At-Risk	.06	1.06	.08	-.49	.61	.30	-.35	.71	.30
Years				-.04	.96	.04	-.05	.95	.08
ELP Score				-.19	.82	.23	-.05	.96	.23
Deviance (Value/df)		1.43		.98			1.01		
Likelihood Ratio χ^2		41.25*		68.28*			13.05		

Note. Am = American; ELA = English language arts; SES = Socioeconomic status; HI = Hawaiian; ELP = English language proficiency; df = Degrees of freedom; Exp = Exponential function; SE = Standard error.

**p* < .01

Table 4.4. History Total Response Times (Minutes) for Forms 1 and 2

			RT (Min)	Exp(Ln(RT)) (Min)	
Form	Student Group	N	Median	Mean	SD
1	Non-EL	533	13.25	14.26	8.00
	High-Prof EL	441	26.00	30.88	21.31
	Mid-Prof EL	295	28.83	34.01	22.11
2	Non-EL	489	13.50	13.98	7.43
	High-Prof EL	479	24.50	30.24	20.59
	Mid-Prof EL	284	38.85	32.62	18.42
	All Students	2521	18.90	24.53	18.82

Note. Prof = Proficient; RT = Response Time; Min = Minutes; Exp = Exponential function; SD = Standard deviation

Table 4.5. History Item-Level Response Times (Seconds) For Forms 1A and 2A

		RT (Sec)	Exp(Ln(RT)) (Sec)	
Form	Student Group	Median	Mean	SD
1A	Non-EL	27.0	25.79	2.25
	High-Prof EL	45.0	47.47	2.56
	Mid-Prof EL	51.5	52.46	2.59
2A	Non-EL	23.0	22.42	2.25
	High-Prof EL	39.0	41.26	2.64
	Mid-Prof EL	46.0	44.70	2.66

Note. Prof = Proficient; RT = Response Time; Sec = Seconds; Exp = Exponential function; SD = Standard deviation

Table 4.6. History ANOVA Summary Table for Response Time Differences

Source	SS	df	MS	F	*p*-value	η^2
Student Group	14.044	2	7.022	56.880	.000	.441
Form	.772	1	.772	6.255	.014	.024
Group*Form	.006	2	.003	.023	.978	.000
Error	17.036	138	.123			
Total	1910.814	144				

Note. SS = Sums of squares; df = Degrees of freedom; MS = Mean square

Table 4.7. History Standardized Mean Differences on Item-Level Response Time

Item	# Accoms	Accom v. Non-Accom (Same Groups)			Accom Items – Group Differences			NonAccom Items – Group Differences		
		NonEL	HighEL	MidEL	NonEL-HighEL	NonEL-MidEL	HighEL-MidEL	NonEL-HighEL	NonEL-MidEL	HighEL-MidEL
11	11	.202*	.200*	.299*	-.653*	-.858*	-.144	-.636*	-.631*	-.034
6	9	.150	.132	.189	-.642*	-.850*	-.165	-.617*	-.739*	-.115
9	9	.252*	.109	.258*	-.817*	-1.085*	-.273*	-.853*	-.992*	-.098
2	8	.421*	.413*	.394*	-.837*	-1.053*	-.188	-.764*	-1.027*	-.229*
15	8	.105	.166	.318*	-.792*	-1.009*	-.193	-.726*	-.728*	-.028
18	7	.245*	.234*	.294*	-.690*	-.689*	-.012	-.613*	-.590*	.043
5	6	.368*	.343*	.353*	-.733*	-.938*	-.188	-.652*	-.776*	-.140
12	6	.155	.263*	.038	-.785*	-.637*	.120	-.639*	-.740*	-.100
19	6	.048	.122	.120	-.693*	-.770*	-.084	-.613*	-.639*	-.071
22	6	.229*	.082	.112	-.537*	-.531*	.001	-.630*	-.618*	.028
16	5	.375*	.208*	.072	-.676*	-.562*	.065	-.732*	-.808*	-.064
21	5	.067	.005	.269*	-.593*	-.834*	-.189	-.650*	-.552*	.076
4	4	.168	.150	.103	-.587*	-.683*	-.096	-.575*	-.744*	-.150
25	4	.111	.110	.146	-.903*	-1.092*	-.199	-.876*	-1.007*	-.148
7	3	.262*	.093	.107	-.523*	-.714*	-.165	-.684*	-.813*	-.139
8	3	.139	.158	.213*	-.778*	-.941*	-.157	-.693*	-.848*	-.113
17	3	-.058	.036	.063	-.639*	-.743*	-.070	-.557*	-.547*	-.036
13	2	-.029	-.030	.013	-.562*	-.637*	-.092	-.577*	-.621*	-.049
20	2	.118	.161	.132	-.834*	-.823*	-.015	-.777*	-.822*	-.042
23	2	.116	-.027	.059	-.707*	-.822*	-.065	-.815*	-.769*	.022
24	2	.215*	.090	.215*	-.806*	-.912*	-.147	-.827*	-.898*	-.020
1	1	.075	.092	.204*	-.537*	-.829*	-.297*	-.470*	-.666*	-.171
10	1	.156	.016	-.098	-.830*	-.798*	-.015	-.894*	-1.009*	-.130
3[a]	0	.045	.025	.025	-.730*	-.835*	-.114	-.745*	-.851*	-.116
14[a]	0	.039	.042	-.018	-.710*	-.855*	-.124	-.700*	-.872*	-.182

Note. Accoms = Accommodations; Accom = Accommodated; HighEL = High-Proficient ELs; MidEL = Mid-Proficient ELs.
[a]No accommodation on either form - differences are simply between Forms 1 and 2.
*Practical significance, standardized mean difference ≥ .20.

Table 4.8. History Items with Practically Significant Response Time Differences Between Forms 1A and 2A

Item #	# Accoms	Non-EL	High-Prof EL	Mid-Prof EL	Total
11	11	●	●	●	3
2	8	●	●	●	3
18	7	●	●	●	3
5	6	●	●	●	3
16	5	●	●		2
9	9	●		●	2
24	2	●		●	2
22	6	●			1
7	3	●			1
12	6		●		1
15	8			●	1
21	5			●	1
8	3			●	1
1	1			●	1
Total		9	6	10	

Table 4.9. History Regression Coefficients Predicting Response Time

		Non-EL			High-Proficient EL			Mid-Proficient EL		
Form	Predictor	B	β	SE	B	β	SE	B	β	SE
1A	Intercept	3.38**		.53	5.96**	-.10	.51	5.07**		.51
	Sex	-.18**	-.14	.04	-.13**	.07	.04	-.11*	-.09	.05
	Hispanic	.42**	.32	.05	-.49**	-.09	.12	.02	.01	.17
	Asian	.43**	.13	.11	-.05	-.003	.25	.01	.001	.37
	Black	.26*	.08	.12	.12	-.34	.13	.60**	.34	.18
	Am Indian	.04	.003	.36	-.74*	-.01	.30	.19	.02	.44
	SES	-.02	-.01	.05	.14*	-.08	.07	.14	.08	.08
	At-Risk	-.05	-.04	.05	-.37*	-.07	.16	-.05	-.01	.17
	ELA Score	.00	-.02	.00	.00*	-.06	.00	.00	-.09	.00
	Native HI	.68	.05	.44	-.01	.07	.15	.47*	.15	.21
	Years				-.04	-.05	.03	.04	-.01	.04
	ELP Score				-.18	-.10	.13	.14	-.06	.14
	R	.329**			.445**			.372**		
	R^2	.108			.198			.138		
2A	Intercept	2.99**		.50	6.05**		.49	4.67**		.51
	Sex	-.16**	-.13	.04	-.10*	-.08	.04	-.07	-.06	.05
	Hispanic	.41**	.33	.05	-.43**	-.31	.11	-.07	-.05	.17
	Asian	.37**	.12	.11	-.13	-.02	.24	-.01	-.001	.38
	Black	.23*	.07	.11	.09	.06	.12	.42*	.24	.18
	SES	-.03	-.02	.04	.13*	.07	.06	.08	.04	.08
	ELA Score	.00	-.01	.00	.00*	-.08	.00	.00	-.07	.00
	ELP Score				-.30*	-.08	.13	-.15	-.05	.14
	Am Indian	.15	.02	.34	-.56	-.07	.29	-.14	-.02	.45
	Native HI	.56	.04	.42	-.08	-.03	.15	.39	.13	.21
	At-Risk	-.03	-.03	.05	-.24	-.05	.15	.06	.14	.17
	Years				-.03	-.05	.02	-.01	-.01	.04
	R	.328**			.407**			.315**		
	R^2	.107			.166			.099		

Note. HI = Hawaiian; Am = American; SES = Socioeconomic status; ELA = English language arts; ELP = English language proficiency; SE = Standard error.
* $p < .05$. ** $p < .01$

Table 4.10. History Correlation Between Test Raw Score and Response Time

				vs. High-Prof EL		vs. Mid-Prof EL	
Form	Student Group	N	R	*z*	R^2_{Diff}	*z*	R^2_{Diff}
1	Non-EL	533	.335**	3.32**	.095	2.44*	.084
	High-Prof EL	441	.133*	--	--	.49	.011
	Mid-Prof EL	295	.169*	--	--	--	--
2	Non-EL	489	.370**	5.01**	.133	3.76**	.126
	High-Prof EL	479	.065	--	--	.55	.007
	Mid-Prof EL	284	.106	--	--	--	--

Note. Prof = Proficient.
*$p < .01$. **$p \leq .001$

Table 4.11. History Accommodation Use Analysis Summary Table

Accommodation Use	Significant	Effect Size
Student Group Differences	All test items*	V Range = .124 to.264
Items with Practical Significance	Items 2, 6, 11, 12, 18, 19, 20, 22-25	V Range = .210 to .264
Predictors		
Non-EL	Full Model*	
	Sex (-)*	Males 20% less
	Hispanic (+)*	Hispanic 34% more
	Am Indian (-)*	Am Indian 87% less
	ELA Score (-)*	<1% less per increase in ELA score
HighProf EL	Full Model*	
	Sex (-)*	Males 22% less
	Hispanic (-)*	Hispanic 41% less
	SES (+)*	Low-SES 50% more

Note. Am = American; ELA = English Language Arts; SES = Socioeconomic status; (-) = negative regression coefficient; (+) = positive regression coefficient.
*$p < .01$

Table 4.12. History Response Time Analysis Summary Table

Test Level Differences	Significant		Effect Size	
Student Group	All groups**		η^2 = .441	
	Non-EL vs. HighProf ELs**		*d* = 1.457	
	Non-EL vs. MidProf ELs**		*d* = 1.660	
Test Form	Accom vs. No Accom*		η^2 = .024	
Accom Items v. Non-Accom Items				
Non-EL	Items 2, 5, 7, 9, 11, 16, 18, 22, 24		*d* = .202 to .421	
HighProf EL	Items 2, 5, 11, 12, 16, 18		*d* = .200 to .413	
MidProf EL	Items 1, 2, 5, 8, 9, 11, 15, 18, 21, 24		*d* = .204 to .394	
Item-Level Differences	**Form 1A**	**Form 2A**	**Form 1A**	**Form 2A**
Non-EL v. HighProf EL	All items	All items	*d* = .537 to .903	*d* = .470 to .894
Non-EL v. MidProf EL	All items	All items	*d* = .531 to 1.092	*d* = .547 to 1.027
HighProf EL v. MidProf EL	Items 1, 9	Item 2	*d* = .297; .273	*d* = .229
Predictors	**Form 1A**	**Form 2A**	**Form 1A**	**Form 2A**
Non-EL	Full Model**	Full Model**	R^2 = .108	R^2 = .107
	Sex (-)**	Sex (-)**		
	Black (+)*	Black (+)*		
	Hispanic (+)**	Hispanic (+)**		
	Asian (+)**	Asian (+)**		
HighProf EL	Full Model**	Full Model**	R^2 = .198	R^2 = .166
	Sex (-)**	Sex (-)*		
	Hispanic (-)**	Hispanic (-)**		
	Am Indian (-)*	SES (+)*		
	SES (+)*	ELP Score (-)*		
	At-Risk (-)*	ELA Score (-)*		
	ELA Score (-)*			
MidProf EL	Full Model**	Full Model**	R^2 = .138	R^2 = .099
	Black (+)**	Black (+)*		
	Native HI (+)*			
	Sex (-)*			
Response Time and Raw Score	**Form 1**	**Form 2**	**Form 1**	**Form 2**
Non-ELs	Correlation**	Correlation**	R^2 = .112	R^2 = .137
	Vs. High-Prof ELs **	Vs. High-Prof ELs**	R^2_{Diff} = .095	R^2_{Diff} = .132
	Vs. Mid-Prof ELs*	Vs. Mid-Prof ELs**	R^2_{Diff} = .084	R^2_{Diff} = .126
HighProf ELs	Correlation**		R^2 = .018	
MidProf ELs	Correlation**		R^2 = .027	

Note. Accom = Accommodated; HI = Hawaiian; Am = American; SES = Socioeconomic status; ELA = English Language Arts; ELP = English Language Proficiency; (-) = negative regression coefficient; (+) = positive regression coefficient.
**p* < .05. ***p* < .01

Table 4.13. Mathematics Test Score Descriptive Statistics

Form	Student Group	N	Mean	SD
1	Non-EL	350	11.29	3.65
	High-Prof EL	472	8.32	3.09
	Mid-Prof EL	289	6.99	2.63
2	Non-EL	316	11.63	3.47
	High-Prof EL	450	8.33	3.29
	Mid-Prof EL	275	7.16	2.81
	All Students	2152	8.96	3.63

Note. Prof = Proficient; SD = Standard deviation

Table 4.14. Mathematics Differences in Accommodation Use Between Student Groups

Item[a]	# Accoms	χ^2	df	Cramer's V
20	1	105.85[b]	2	.309
18	6	112.86	12	.225
5	8	99.49	16	.219
12	5	105.27	10	.218
24	3	96.95	6	.209
25	8	88.79	16	.207
4	6	89.19	12	.200
9	3	78.97	6	.195
7	6	73.24	12	.188
19	4	72.46	8	.187
15	4	69.87	8	.183
6	6	73.35	12	.182
3	3	67.67	6	.180
16	4	70.57	8	.178
13	4	60.88	8	.171
10	4	60.26	8	.165
22	8	58.75	16	.163
1	2	34.84	4	.129
2	2	7.40	4	.058

Note. Accoms = # of accommodations available for the item; df = Degrees of freedom.
[a]Items 8, 11, 14, 17, 21, and 23 did not have test accommodations and are not included in the table
[b]All χ^2 were significant ($p < .001$), except Item 2

Table 4.15. Mathematics Negative Binomial Regression Coefficients Predicting Accommodation Use

	Non-EL			High-Proficient EL			Mid-Proficient EL		
	B	Exp(B)	SE	B	Exp(B)	SE	B	Exp(B)	SE
Intercept	4.90**	133.71	1.10	3.98**	53.36	.94	1.88*	6.54	.85
Sex	.18*	1.20	.09	-.09	.91	.08	-.18	.84	.10
SES	.04	1.04	.10	.23*	1.26	.11	.08	1.08	.15
At-Risk	.01	1.01	.10	-.07	.93	.39	1.44**	4.23	.46
Black	-.08	.92	.11	-.02	.98	.22	.83**	2.30	.30
Native HI				.00	1.00	.27	1.05**	2.84	.38
Hispanic	.04	1.04	.11	-.34	.71	.20	.64*	1.89	.28
Am Indian	-.92	.40	1.16	-.13	.88	.59	-.01	.99	.60
Asian	-.17	.84	.33	-.47	.62	.47	-1.00	.37	.90
ELA Score	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
Years				.03	1.03	.04	.02	1.02	.08
ELP Score				-.22	.80	.24	-.42	.65	.24
Deviance (Value/df)		1.56		1.41			1.37		
Likelihood Ratio χ^2		16.17*		27.26**			25.54**		

Note. HI = Hawaiian; SES = Socioeconomic status; Am = American; ELA = English language arts; ELP = English language proficiency; df = Degrees of freedom; Exp = Exponential function; SE = Standard error.

$*p < .05$. $**p < .01$

Table 4.16. Mathematics Total Response Times (Minutes) for Forms 1 and 2

			RT (Min)	Exp(Ln(RT)) (Min)	
Form	Student Group	N	Median	Mean	SD
1	Non-EL	350	29.86	31.70	15.66
	High-Prof EL	471	52.70	55.44	29.43
	Mid-Prof EL	289	53.98	57.33	28.52
2	Non-EL	316	31.51	32.66	15.41
	High-Prof EL	447	53.70	57.71	28.80
	Mid-Prof EL	275	53.32	57.85	28.92
	All Students	2148[a]	44.93	49.26	28.02

Note. Prof = Proficient; RT = Response Time; Min = Minutes; Exp = Exponential function; SD = Standard deviation

[a]Response time missing for 4 students

Table 4.17. Mathematics Item-Level Response Times (Seconds) For Forms 1A and 2A

		RT (Sec)	Exp(Ln(RT)) (Sec)	
Form	Student Group	Median	Mean	SD
1A	Non-EL	63.00	50.43	2.75
	High-Prof EL	106.00	86.88	3.06
	Mid-Prof EL	112.00	91.82	2.90
2A	Non-EL	55.00	42.39	2.87
	High-Prof EL	92.00	70.66	3.23
	Mid-Prof EL	95.00	73.05	3.11

Note. Prof = Proficient; RT = Response Time; Sec = Seconds; Exp = Exponential function; SD = Standard deviation

Table 4.18. Mathematics ANOVA Summary Table for Response Time Differences

Source	SS	df	MS	F	*p*-value	η^2
Student Group	8.741	2	4.371	16.490	.000	.201
Form	1.334	1	1.334	5.035	.027	.031
Group*Form	.016	2	.008	.031	.969	.000
Error	33.396	126				
Total	2356.875	132				

Note. SS = Sums of squares; df = Degrees of freedom; MS = Mean square

Table 4.19. Mathematics Standardized Mean Differences on Item-Level Response Time

		Accom v. Non-Accom (Same Groups)			Accom Items – Group Differences			NonAccom Items – Group Differences		
Item	# Accoms	NonEL	HighEL	MidEL	NonEL-HighEL	NonEL-MidEL	HighEL-MidEL	NonEL-HighEL	NonEL-MidEL	HighEL-MidEL
5	8	.227*	.220*	.269*	-.822*	-.877*	-.076	-.727*	-.789*	-.027
22	8	.047	-.132	-.025	-.504*	-.630*	-.115	-.714*	-.725*	-.012
25	8	.123	.202*	.173	-.717*	-.716*	-.029	-.563*	-.659*	-.057
4	6	-.064	-.003	.008	-.386*	-.419*	-.008	-.354*	-.355*	.003
6	6	.041	.058	.187	-.436*	-.683*	-.167	-.433*	-.501*	-.058
7	6	.054	.186	-.028	-.470*	-.474*	.002	-.281*	-.547*	-.214*
18	6	.039	-.107	-.002	-.378*	-.489*	-.089	-.577*	-.537*	.016
12	5	-.080	-.054	.124	-.428*	-.573*	-.105	-.451*	-.366*	.070
10	4	.036	-.010	.063	-.403*	-.507*	-.085	-.485*	-.522*	-.019
13	4	.049	.104	.142	-.540*	-.484*	.073	-.460*	-.364*	.097
15	4	.108	.184	.130	-.603*	-.531*	.061	-.477*	-.480*	.004
16	4	.092	-.014	-.020	-.438*	-.508*	-.046	-.551*	-.613*	-.054
19	4	.180	.206*	.133	-.498*	-.424*	.076	-.440*	-.460*	-.006
3	3	.061	.159	.115	-.483*	-.467*	.017	-.346*	-.400*	-.034
9	3	.090	.113	-.010	-.327*	-.332*	.003	-.263*	-.390*	-.117
24	3	.051	-.098	-.060	-.717*	-.737*	-.035	-.897*	-.890*	.000
1	2	.057	.023	.138	-.807*	-1.015*	-.164	-.761*	-.899*	-.053
2	2	.134	-.105	.065	-.437*	-.671*	-.157	-.693*	-.671*	.014
20	1	-.032	-.108	.014	-.525*	-.711*	-.181	-.670*	-.710*	-.077
8[a]	0	.040	-.038	.054	.110	-.074	-.165	.030	-.049	-.071
11[a]	0	-.051	-.062	.069	-.394*	-.470*	-.044	-.456*	-.346*	.088
14[a]	0	.094	-.072	.004	.044	-.121	-.152	-.119	-.203*	-.074
17[a]	0	.004	-.176	-.146	-.438*	-.449*	.015	-.653*	-.600*	.061
21[a]	0	.002	-.103	-.020	-.481*	-.569*	-.072	-.627*	-.604*	.010
23[a]	0	.053	-.064	.041	-.440*	-.422*	.018	-.601*	-.450*	.129

Note. Accoms = Accommodations; Accom = Accommodated; HighEL = High-Proficient ELs; MidEL = Mid-Proficient ELs..
[a]No accommodation on either form - differences are simply between Forms 1 and 2.
*Practical significance, standardized mean difference ≥ .20.

Table 4.20. Mathematics Items with Practically Significant Response Time Differences between Forms 1A and 2A

Item #	# Accoms	Non-EL	High-Prof EL	Mid-Prof EL	Total
5	8	●	●	●	3
25	8		●		1
19	4		●		1
Total		1	3	1	

Table 4.21. Mathematics Regression Coefficients Predicting Response Time

		Non-EL			High-Proficient EL			Mid-Proficient EL		
Form	Predictor	B	β	SE	B	β	SE	B	β	SE
1A	Intercept	3.44**		.70	7.15**		.70	5.54**		.58
	Black	.20**	.12	.07	.34*	.16	.17	.60**	.29	.20
	Sex	-.16**	-.11	.06	-.03	-.02	.06	-.17*	-.11	.07
	ELP Score				-.65**	-.13	.18	-.09	-.02	.17
	Native HI				.37	.09	.21	.84**	.20	.26
	Hispanic	.08	.05	.08	-.30	-.16	.16	.12	.07	.18
	Am Indian	.34	.02	.71	.05	.01	.41	.07	.01	.41
	Asian	.04	.01	.22	-.18	-.02	.35	-.47	-.04	.54
	SES	.07	.04	.07	-.004	-.002	.09	.02	.01	.10
	At-Risk	-.10	-.06	.06	-.18	-.03	.26	-.17	-.03	.30
	ELA Score	.00	.03	.00	.00	-.02	.00	.00	-.08	.00
	Years				-.05	-.05	.03	-.02	-.02	.06
	R	.178**			.355**			.319**		
	R^2	.032			.126			.102		
2A	Intercept	3.41**		.65	6.69**		.62	4.88**		.52
	Sex	-.15**	-.11	.05	-.04	-.03	.05	-.19**	-.14	.06
	Black	.22**	.14	.07	.17	.09	.15	.60**	.32	.18
	ELP Score				-.61**	-.14	.16	-.17	-.05	.15
	Hispanic	.05	.03	.07	-.34*	-.20	.14	.14	.09	.17
	Native HI				.21	.06	.19	.72**	.18	.24
	Asian	.001	.00	.20	-.40	-.05	.31	-1.08*	-.10	.49
	Am Indian	.26	.02	.66	-.22	-.02	.36	.21	.03	.37
	SES	.04	.03	.06	-.09	-.04	.08	.08	.04	.09
	At-Risk	-.10	-.07	.06	-.23	-.04	.23	-.06	-.01	.27
	ELA Score	.00	.02	.00	.00	-.01	.00	.00	-.05	.00
	Years				-.04	-.05	.03	.00	.00	.051
	R	.192**			.333**			.344**		
	R^2	.037			.111			.118		

Note. HI = Hawaiian; ELP = English language proficiency; Am = American; SES = Socioeconomic status; ELA = English language arts; SE = Standard error.
**p* < .05. ***p* < .01

Table 4.22. Mathematics Correlation Between Test Raw Score and Response Time

Form	Student Group	N	R	vs. High-Prof EL		vs. Mid-Prof EL	
				z	R^2_{Diff}	*z*	R^2_{Diff}
1	Non-EL	350	.433**	3.01**	.127	3.85**	.163
	High-Prof EL	471	.245**	--	--	1.25	.036
	Mid-Prof EL	289	.155*	--	--	--	--
2	Non-EL	316	.344**	1.63	.064	3.41**	.113
	High-Prof EL	447	.234**	--	--	2.11*	.049
	Mid-Prof EL	275	.076	--	--	--	--

Note. Prof = Proficient.
Effect size = Difference between squared correlations.
*$p < .05$. **$p \leq .001$

Table 4.23. Mathematics Accommodation Use Analysis Summary Table

Accommodation Use	Significant	Effect Size
Student Group Differences	All test items (except Item 2)**	V Range = .129 to.309
Items with Practical Significance	Items 4, 5, 12, 18, 20, 24, 25	V Range = .200 to .309
Predictors		
Non-EL	Full Model**	
	Sex (+)**	Males 20% more
HighProf EL	Full Model**	
	SES (+)*	Low-SES 25% more
Mid-Prof EL	Full Model**	
	At-Risk (+)**	At-Risk 323% more
	Black (+)**	Black 130% more
	Native HI (+)**	Native HI 184% more
	Hispanic (+)*	Hispanic 89% more

Note. HI = Hawaiian; SES = Socioeconomic status; (-) = negative regression coefficient; (+) = positive regression coefficient.
*$p < .05$. **$p < .01$

Table 4.24. Mathematics Response Time Analysis Summary Table

Test Level Differences	Significant		Effect Size	
Student Group	All groups**		η^2 = .201	
	Non-EL vs. HighProf ELs**		*d* = .848	
	Non-EL vs. MidProf ELs**		*d* = .918	
Test Form	Accom vs. No Accom*		η^2 = .014	
Accom Items v. Non-Accom Items				
Non-EL	Item 5		*d* = .227	
HighProf EL	Items 5, 19, 25		*d* = .220; .206; .202	
MidProf EL	Item 5		*d* = .269	
Item-Level Differences	**Form 1A**	**Form 2A**	**Form 1A**	**Form 2A**
Non-EL v. HighProf EL	All items (except 8, 14)	All items (except 8, 14)	*d* = .327 to .882	*d* = .263 to .897
Non-EL v. MidProf EL	All items (except 8, 14)	All items (except 8)	*d* = .332 to 1.015	*d* = .203 to .899
HighProf EL v. MidProf EL	None	Item 7		*d* = .214
Predictors	**Form 1A**	**Form 2A**	**Form 1A**	**Form 2A**
Non-EL	Full Model** Black (+)** Sex (-)**	Full Model** Black (+)** Sex (-)**	R^2 = .032	R^2 = .037
HighProf EL	Full Model** Black (+)* ELP Score (-)**	Full Model** Hispanic (-)* ELP Score (-)**	R^2 = .126	R^2 = .111
MidProf EL	Full Model** Black (+)** Sex (-)* Native HI (+)**	Full Model** Black (+)** Sex (-)** Native HI (+)** Asian (-)*	R^2 = .102	R^2 = .118
Response Time and Raw Score	**Form 1**	**Form 2**	**Form 1**	**Form 2**
Non-ELs	Correlation** Vs. High-Prof ELs ** Vs. Mid-Prof ELs*	Correlation** Vs. Mid-Prof ELs**	R^2 = .187 R^2_{Diff} = .127 R^2_{Diff} = .163	R^2 = .118 R^2_{Diff} = .113
HighProf ELs	Correlation**	Correlation** Vs. Mid-Prof ELs*	R^2 = .060	R^2 = .055 R^2_{Diff} = .049
MidProf ELs	Correlation*		R^2 = .024	

Note. Accom = Accommodated; HI = Hawaiian; ELP = English Language Proficiency; (-) = negative regression coefficient; (+) = positive regression coefficient.
*p < .05. **p < .01

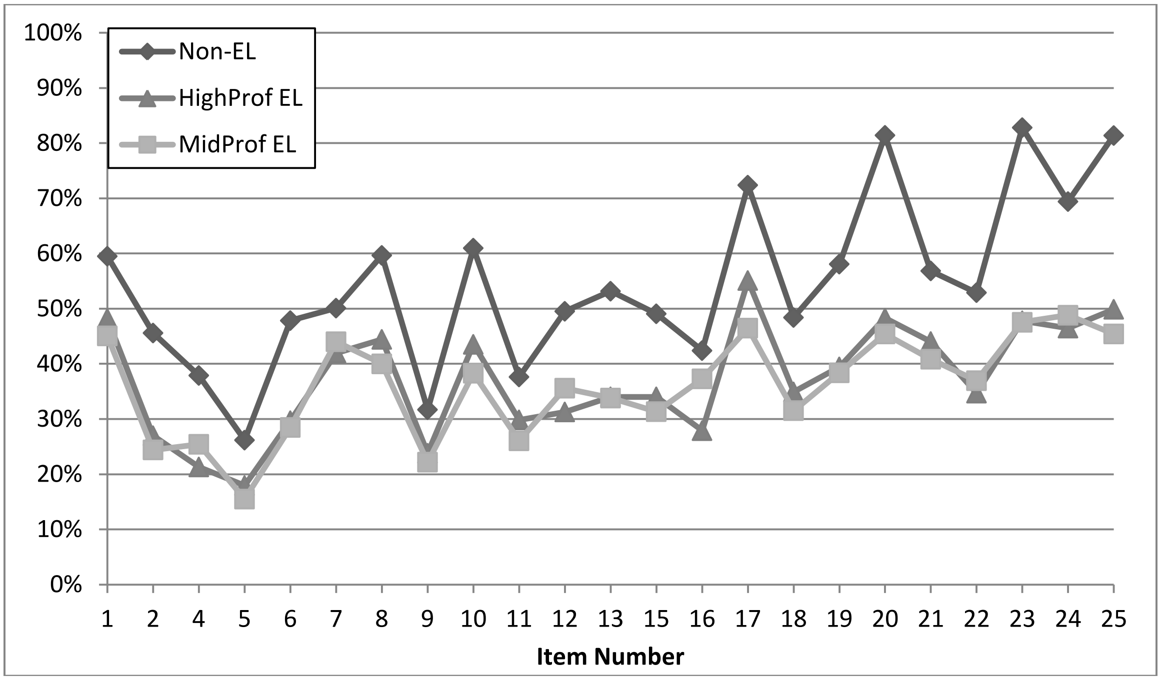


Figure 4.1. History Percent of Students Not Using an Accommodation (0 Use)

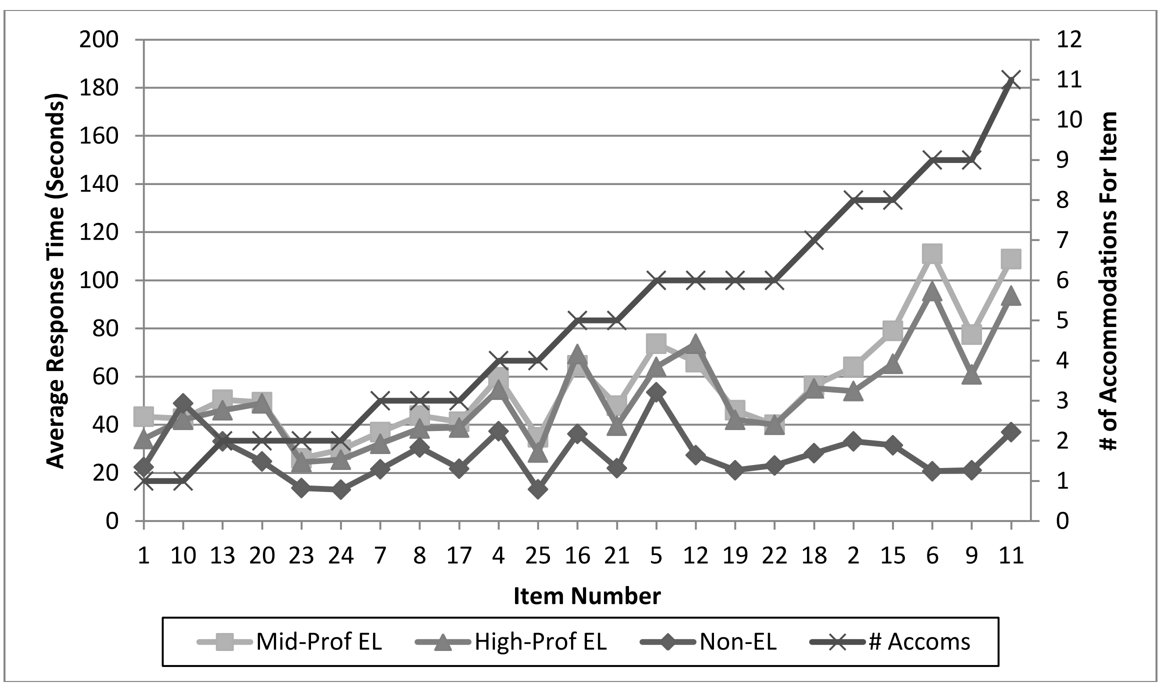


Figure 4.2. History Relationship Between Number of Accommodations on an Item and Average Item Response Time.

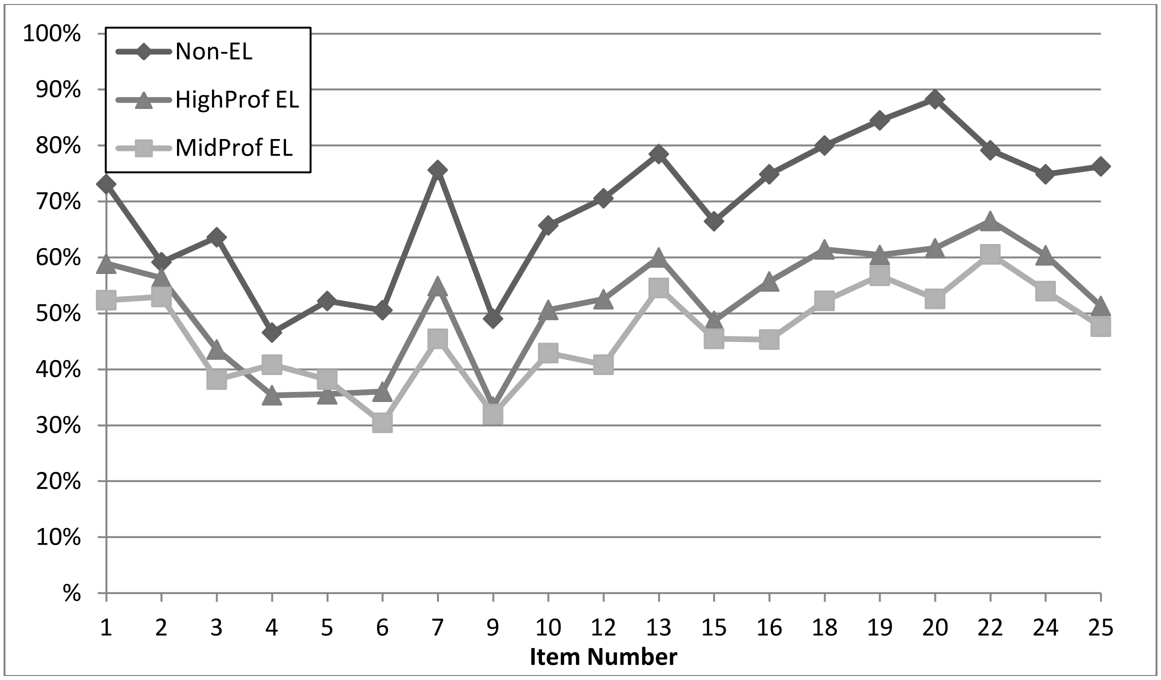


Figure 4.3. Mathematics Percent of Students Not Using an Accommodation (0 Use)

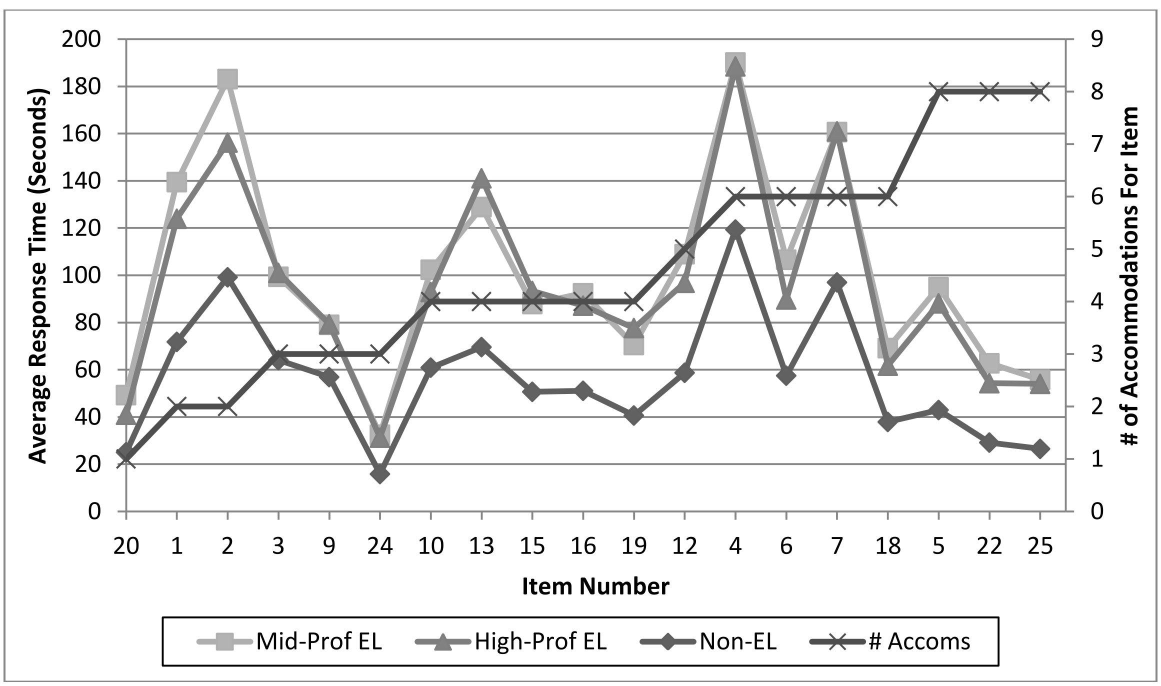


Figure 4.4. Mathematics Relationship Between Number of Accommodations on an Item and Average Item Response Time.

CHAPTER 5

DISCUSSION

5.1 Overview

The purpose of this study was to evaluate the validity of computer-based accommodations for ELs on History and Math assessments using evidence based on response processes, specifically accommodation use and response time analysis. As computer-based testing increases throughout K-12 assessment, there is a need to evaluate the effectiveness and validity of computer-based accommodations for ELs. Fortunately, the use of technology provides additional features to evaluate the validity of test accommodations and provides more complete evidence on interpretations of test scores for ELs. Specifically, evidence based on response processes can show how accommodations may be benefiting ELs, supporting accommodation use, improving test validity, and improving academic instruction (Scarpati et al., 2011). Because of the limited research on computer-based accommodations for ELs, there is also limited research on obtaining different sources of validity evidence to evaluate accommodations. Ultimately, this study attempted to fill the void in the current investigation on test accommodations for ELs by gathering new evidence that can better inform the validity of test scores for ELs through examination of the following research questions:

1. Do ELs use accommodations significantly more often than non-ELs?
2. What characteristics of ELs and non-ELs predict accommodation use?
3. Do ELs and non-ELs with accommodations take significantly longer to complete items than ELs and non-ELs without accommodations?

4. What characteristics of ELs and non-ELs predict response time on accommodated and non-accommodated test items?
5. What is the relationship between student proficiency, response time, and EL and non-EL accommodation status?

This chapter discusses the results of the above research questions in detail. This chapter will be structured by first discussing accommodation use results, followed by response time results, and results discussing the combination of accommodation use and response time. Each section will begin with a general overview of the results for History and Math, followed by a discussion of differences in results across the two subject areas. There will be a discussion about what these results might mean and implications of these results. After discussion of results, discussion will be made about how the results impact the validity of the test accommodations, limitations of the study, and directions for future research. This chapter will conclude with an overall conclusion of the study.

5.2 Accommodation Use Results

Due to lack of research on computer-based accommodations, there has also been a lack of research investigating whether ELs are actually using the test accommodations that are being provided to them. If students are not using the accommodation, then the accommodation will fail to provide the linguistic support that an EL might need to ultimately level the playing field. Research on paper-and-pencil assessments indicated that few students used the customized English dictionary provided as an accommodation to ELs (Abedi et al., 2003b). This lack of use could have been due to students not finding the accommodation useful within the first few attempts of the using the accommodation. Abedi et al. (2003b) also examined how often students looked up words with the pop-up glossary accommodation, finding that ELs glossed over twice as many words as non-ELs at Grade 8, suggesting that ELs at Grade 8 were effectively using the accommodation.

In this section discussion is made about whether the findings were consistent to previous research on accommodation use, and how the results might inform the effectiveness and validity of test accommodations. Research questions in relation to accommodation use investigated differences in accommodation use across student group and characteristics predicting accommodation use.

5.2.1 Differences in Accommodation Use

Statistically significant differences in accommodation use were found between non-EL and EL student groups on both the History and Mathematics assessments. On History, statistically significant differences on the total test form were only found between non-ELs and EL student groups, and not between the two EL groups. Math results, however, did reveal statistically significant differences on total accommodation use between the two EL student groups. Item-level frequencies of accommodation use were also investigated for each student group. In both subject areas, results showed that non-ELs were least likely to access accommodations. Many of the items showed non-ELs either accessing zero accommodations, or only accessing one accommodation. It is likely that non-ELs accessed one accommodation simply out of curiosity, found that it was not helpful in improving their understanding of the test item, and therefore did not continue to use available accommodations on that item. For the EL student groups, however, it was very common to see large percentages of students accessing almost all of the accommodations available on that item.

To further examine differences in accommodation use, statistical item-level analyses were also conducted. Of the 23 accommodated items provided on the History assessment, all items showed statistically significant differences in accommodation use among the three student groups ($p < .001$). When examining the effect size of these differences, 11 items showed moderate associations (Cramer's V $\geq .20$) between accommodation use and student group. The 11 items with moderate associations were typically towards the end of the assessment, and were items that showed large differences in item difficulty between the three student groups. Because of this interesting result, accommodation use and whether it was linked to item location was examined. Results indicated that accommodation use declined as students progressed throughout the assessment.

Results for item-level differences in accommodation use on the Mathematics assessment yielded similar results to the History assessment. Of the 19 accommodated items, 18 items showed statistically significant differences in accommodation use ($p < .001$). A total of 7 items yielded moderate associations between accommodation use and student group. Similar to History these items were typically towards the end of the assessment, and were typically difficult items across the three groups. Additionally, the trend of decreased accommodation use as the assessment progressed was also found.

Abedi et al. (2003b) found that lack of accommodation use could be due to students not finding the accommodation useful within the first few attempts of using the accommodation. It is possible that this same phenomenon was occurring on both the History and Math assessments. Accommodation use declined for all student groups as the assessment progressed, suggesting that students were not finding the accommodation useful at the beginning of the assessment, and therefore stopped accessing the accommodations. This decrease in accommodation use was largest for non-ELs on both subject areas, which was expected. Since the accommodations were not intended for non-ELs and were intended to remove the linguistic barrier for ELs, a decline would be expected in accommodation use for those students. On the last ten items of the History assessment, between 40-80% of non-ELs stopped access accommodations, compared to 30-45% of ELs. Similarly, on the last ten items of the math assessment, 60-90% of non-ELs stopped using accommodations compared to 45-65% of ELs.

Understanding whether students are using accommodations is essential for understanding how accommodations are directly impacting student test scores. On the History assessment, an average of 55% of non-ELs, 37% of high-proficient ELs, and 36% of mid-proficient ELs did not access accommodations on the test items. On Math these percentages were even higher, with 69% of non-ELs on average not using accommodations on a single item, 52% of high-proficient ELs, and 46% of mid-proficient ELs. These results provide interesting insight into past research on test accommodations. Previous accommodation research has focused on the interaction hypothesis, improving scores for ELs and not for non-ELs (see Scarpati et al., 2011; Sireci et al., 2003). If high percentages of students are not using the accommodations provided, any test scores for students taking the test with accommodations are also likely to be affected, thus impacting results typically found using the interaction hypothesis method. Specifically, since many non-ELs were not accessing the accommodations, no increase in test scores is expected. Similarly, if scores improve drastically for ELs with only a small percentage of students accessing the accommodations, it could be that the accommodations are only improving scores of very specific students, rather than all ELs. Based on these results, future studies should consider focusing on students actually using the accommodations prior to investigating the interaction hypothesis. This method would provide a more accurate interpretation of accommodation effectiveness for ELs.

In relation to the differential boost hypothesis, the idea that ELs will benefit more than non-ELs when provided with the same accommodation (Cho et al., 2012), results suggest that ELs are in fact using accommodations more often than non-ELs. Knowing that ELs are using accommodations more than non-ELs could suggest that ELs are benefiting more from receiving accommodations. Again, however, comparisons of performance between non-ELs and ELs actually using the accommodations are suggested for future research.

5.2.2 Characteristics Predicting Accommodation Use

In addition to differences in accommodation use, predictors of accommodation use were also investigated for the three student groups. Focusing on History results, statistically significant predictors were found for both the non-EL model and the high-proficient EL model. In both models, males were approximately 20% less likely to access accommodations as compared to females. Interestingly, for non-ELs, Hispanic students were 34% more likely to use accommodations which was opposite of the result for high-proficient ELs where Hispanic students were 41% *less* likely to use accommodations. For non-ELs, American Indian students were 87% less likely to use accommodations, and for every one point increase in ELA score, less than 1% of students were less likely to use the accommodations. For high-proficient ELs, low-SES students were 50% more likely to use accommodations than students with high-SES.

For the Math results, statistically significant predictors were found for all three student group models. For the non-EL model, sex was the only significant predictor, with males 20% more likely to use accommodations as compared to females. This result is the opposite result found on the History assessment. For high-proficient ELs, only SES was a statistically significant predictor, with low-SES students 25% more likely to use accommodations. SES was statistically significant predictor of accommodation use for high-proficient ELs on the History assessment as well. Unlike History, the model for mid-proficient ELs on the Math assessment showed the following statistically significant positive predictors: at-risk, Black, Native Hawaiian, and Hispanic. Table 5.1 shows the comparisons in significant predictors across subject area and student groups.

Previous research on accommodation use has not looked directly at student characteristics. Abedi et al. (2003b) looked at differences in use between Grade 4 and Grade 8, but did not consider other characteristics in relation to accommodation use. One of the more interesting results was found on the History assessment with non-EL Hispanic students more likely to access an accommodation, and high-proficient EL Hispanic students less likely to access an accommodation. This difference could be due to the sample used in the study. As seen in Table 3.1, about 50% of the non-EL population was Hispanic. This was compared to approximately 20% of the non-EL population taking the Math assessment. In comparison, about 70% of high-proficient ELs were identified as Hispanic (see Table 3.1). ELA score differences were investigated between Hispanic non-ELs and high-proficient ELs to see if performances were similar on the statewide reading assessment, which could inform the differences in accommodation use regression results. Results showed Hispanic non-ELs performing statistically significantly higher on the ELA test as compared to Hispanic high-proficient ELs. Even though Hispanic non-ELs still performed higher than Hispanic high-proficient ELs, the non-EL Hispanic students still could have been former ELs, making those students more likely to access the accommodations due to that artifact. It is also important to note that the difference in Hispanic predictor direction for high-proficient ELs could have been due to the impact of other predictors within the model. Although comparisons in predictors are being made across the three student groups, models were still completed separately for each group, meaning all predictors were relative to that particular student group. For example, other factors, such as SES, most likely played a larger role in accommodation use than Hispanic status for the high-proficient EL students.

Another interesting result was found in the non-EL models. On the History assessment, males were 20% less likely to access an accommodation, and on the Math assessment males were 20% *more* likely to access an accommodation. These results could be related to differences in gender performance within the Science, Technology, Engineering, and Mathematics (STEM) fields, and more specifically, the concept of stereotype threat. Stereotype threat is “the threat that members of a stigmatized group experience when they believe that they may, by virtue of their performance in a domain of relevance, confirm a negative stereotype about themselves and members of their group” (Kellow & Jones, 2008, p. 95). In relation to gender differences in the STEM fields, there is a negative stereotype that females do not perform as well as males. According to the theory around stereotype threat, if a female were to internalize this stereotype while taking a math assessment, she would be more likely to underperform as compared to males (Spencer, Steele, & Quinn, 1999). This phenomenon may explain the differences in accommodation use across subject area. It could be that males are more likely to access an accommodation on a math assessment simply because females may have different levels of motivation on the assessment due to the stereotype threat they are experiencing, and are therefore less likely to access the accommodations.

Although different predictors were found across student groups and subject areas, there were some consistent predictors across models and subject areas (see Table 5.1). Specifically, sex was a significant predictor for both non-ELs and high-proficient ELs on the History assessment, and was also a statistically significant predictor for non-ELs on the Math assessment. Similarly, SES was a statistically significant predictor for high-proficient ELs on both subject areas. Lastly, Hispanic was a significant predictor for both non-ELs and high-proficient ELs on the History assessment, and was a significant predictor for mid-proficient ELs on the Math assessment. To understand why certain predictors play a major role in accommodation use, it is helpful to consider the demographics of the overall EL population. For example, it is not surprising that Hispanic status was a significant predictor given that Hispanic students make up 80% of the EL population (Pitoniak et al., 2009). Similarly, the literature on ELs has noted that many ELs tend to be concentrated in public schools, located in central cities in areas with high levels of poverty (Fry, 2008), so it is not surprising that SES plays a large role in EL accommodation use as well.

Understanding which students are likely to use accommodations can assist in interpreting results of studies examining the effectiveness and validity of test accommodations. Previous literature investigating predictors of EL performance found that amount of time living in the United States, student perception of ability, and language instruction were all statistically significant predictors (Abedi et al., 2001; Abedi et al., 2003b; Abedi et al., 2000; Hofstetter, 2003). The only consistent predictor between this study and previous literature was the amount of time living in the United States, which ultimately was not a statistically significant predictor of accommodation use for any of the models. This could have been due to the way at which EL groups were developed. Specifically, high-proficient and mid-proficient EL student groups were developed based on their amount of time living in the United States, and ELP scores. It could be that amount of time living in the United States was too homogenous among the EL groups making it a non-significant predictor across EL models.

Future research in relation to predictors of performance should also examine whether those same predictors of performance are consistent with predictors of accommodation use. The results of this study showed important differences across subject areas that may be important for future research. Because different demographic variables are likely to predict accommodation use differently across subject area, it is important to consider how the subject area could impact the interpretation of results for test accommodation research studies.

5.3 Response Time Results

In addition to offering the ability to collect information on student accommodation use, computer technology also allows for collection of response time information on individual test items and on full tests for all students. Response time information can be integrated into validation processes (Zenisky & Baldwin, 2006), and can assist in understanding whether students take longer to process when items are accommodated. It is likely that ELs will take longer than non-ELs when provided with the same accommodation. This would indicate longer processing time for ELs as compared to non-ELs. Investigating response time differences between subgroups is essential in ensuring equity (Schnipke & Scrams, 2002).

Previous research on response time has indicated that high performing examinees tend to take longer on test items as compared to examinees with lower performance (Chang, et al., 2005). In relation to subgroup differences, research has suggested that ELs take significantly longer on test items as compared to non-ELs (Schnipke & Pashley, 1997; Zenisky & Baldwin, 2006). Similarly, in relation to accommodated test items, Abedi et al. (2003b) noted that at Grade 4, there were no significant differences in response time when using a pop-up glossary between ELs and non-ELs; however, significant differences in response time were found between ELs and non-ELs at Grade 8.

This section will discuss similarities in this study's findings in relation to previous research. Specifically, this section discusses differences in response time across groups, significant predictors of response time, and the relationship between student performance and response time.

5.3.1 Differences in Response Time

Differences in response time were examined between student group and between accommodated and non-accommodated test items. Additionally, the interaction between student group and test form (accommodated vs. non-accommodated) was examined. On both the History and Math assessments, statistically significant main effects were found for student group and for test form, but not for the interaction. Specifically, statistically significant differences in response time were found between non-ELs and both EL groups, but not between mid-proficient and high-proficient ELs. On a given History item, the average response time for accommodated items was 26, 47, and 52 seconds for non-ELs, high-proficient ELs, and mid-proficient ELs, respectively. On the History non-accommodated items, response times were 22, 41, and 45 seconds for the three student groups, respectively. Math items showed longer response times on average as compared to History. Specifically, on accommodated items non-ELs took approximately 50 seconds, high-proficient ELs took approximately 87 seconds, and mid-proficient ELs took approximately 92 seconds. On non-accommodated items average response times included 42, 71, and 73 seconds for non-ELs, high-proficient ELs, and mid-proficient ELs, respectively. Differences in response time across subject area may have been linked to total test score. Across the two subject areas, all student groups had lower test performance on the Math assessment as compared to the History assessment. The Math test could have been more difficult, required more computation, and thus required more processing time per item as compared to History.

Statistically significant differences in response time between non-ELs and ELs were consistent with results from previous literature (Schnipke & Pashley, 1997; Zenisky & Baldwin, 2006; Abedi et al., 2003b). These results were expected given that ELs are likely to take longer to process test items regardless of whether the item was accommodated or not. Literature on the achievement gap between ELs and non-ELs has indicated a gap of around one standard deviation (Kim & Herman, 2009; Galindo, 2009; Ready & Tindal, 2006), suggesting that EL students are struggling with test content even when receiving test accommodations. Because ELs are likely to be struggling with the content, longer response times are expected due to longer processing time.

Statistically significant differences were also found between accommodated and non-accommodated test items, with accommodated test items requiring longer response time. This result was also expected since more time is required to select and examine the accommodations. To further examine this difference, standardized mean differences were examined between accommodated and non-accommodated items within a student group. Standardized mean differences of .20 or greater were considered of practical significance (Cohen, 1988). On History, a total of 9, 6, and 10 items showed practical significant differences in response time between accommodated and non-accommodated items for non-ELs, high-proficient ELs, and mid-proficient ELs, respectively. Four of the items showing practically significant differences were common between the groups and all had 6 or more accommodations. One additional item was common between non-ELs and high-proficient ELs, and two items were common between non-EL and mid-proficient ELs. These results suggest that these accommodations may need to be used in conjunction with an extended time accommodation for future History assessments, if those assessments have a time limit. Because students are requiring more processing time when using the accommodations, without sufficient time to complete the assessment, those students would be more likely to engage in rapid-guessing behavior, and thus could be disadvantaged by receiving the accommodation.

On the Math assessment, however, only a total of 1, 3, and 1 items showed practical significant differences in response time between accommodated and non-accommodated items for non-ELs, high-proficient ELs, and mid-proficient ELs, respectively. The one item for both non-ELs and mid-proficient ELs was common among all three groups and had a total of 8 accommodations available. Although the accommodated items did not require longer processing time, likely due to the difficulty in test content, it is still important to consider the use of extended time with the accommodations for future studies.

It was interesting to find that not all items showed practical significant differences in response time between accommodated and non-accommodated items, especially on the Math assessment. Reasons for the small number of practically significant differences could be that although some students were accessing accommodations, the accommodation did not add processing time. For example, it could have been that EL students were struggling with the content regardless of whether the item was accommodated resulting in similar processing times between the accommodated and non-accommodated test item. These results can aid in interpreting mean differences in performance with and without test accommodations by providing information about student processing time with and without accommodations, and should be considered in future evaluation of test accommodations. Specifically, response time analysis can aid in determining whether extended time may needed in conjunction with the computer-based accommodations, and how that could impact seat time.

5.3.2 Characteristics Predicting Response Time

Investigating factors that predict response time can inform which factors are likely to predict longer processing time for students. Results for History found that the most variance in response time for accommodated items was explained in the high-proficient EL model, followed by the mid-proficient EL model, and non-EL model. Across all three models predicting accommodated item response time, sex was a significant negative predictor meaning males took less time to respond to items as compared to females. Black, Hispanic, and Asian ethnic predictors were also statistically significant for the non-EL model. For the high-proficient EL model, statistically significant ethnic predictors included Hispanic, and American Indian, and other predictors included SES, at-risk, and ELA score. Black and Native Hawaiian ethnic predictors were statistically significant for the mid-proficient EL model. Similar to the results found on the accommodation use models, the Hispanic predictor was in the opposite direction for the non-EL and high-proficient EL models with non-EL Hispanic students taking longer, and high-proficient ELs taking less time (see Table 5.1). These results were expected given that non-EL Hispanic students were also more likely to access accommodations, and high-proficient EL Hispanic students were less likely to access accommodations.

History results predicting non-accommodated item response time found the most variance explained in the high-proficient EL model, followed by the non-EL model and mid-proficient EL model. The same statistically significant predictors that were found on the accommodated item model for non-ELs were also found on the non-accommodated model. For high-proficient ELs, significant predictors included sex, Hispanic, SES, ELP score, and ELA score. For mid-proficient ELs, only Black was a statistically significant predictor of non-accommodated item response time. Again, opposite directions in the Hispanic predictor for non-ELs and high-proficient ELs were found. This was interesting given that these items were not accommodated. As previously noted, about 50% of the non-ELs taking the History assessment were Hispanic which could have impacted these results. It could have been that the non-EL Hispanic students were former EL students and still required longer processing time on test items.

Math results showed a similar trend as History in relation to amount of variance explained in response time with the most variance explained in the high-proficient EL models, followed by the mid-proficient EL model and non-EL model. That being said, the amount of variance explained was lower than History. The two statistically significant predictors for the non-EL models included Black status. Both of these predictors were also statistically significant in the non-EL models on the History assessment (see Table 5.1 for comparisons between subjects). For high-proficient ELs, statistically significant predictors on the accommodated item model included Black and ELP score. For the non-accommodated item model both Hispanic status and ELP score were statistically significant predictors of response time. These two predictors were also statistically significant on the History non-accommodated model for high-proficient ELs. For the mid-proficient accommodated model significant predictors included Black status, sex, and Native Hawaiian. These results were identical to those found on the History assessment. These same three predictors plus Asian status were statistically significant on the Math non-accommodated model for mid-proficient ELs.

Overall History and Math results yielded fairly similar predictors of accommodated and non-accommodated item response time across the three student groups. The most consistent predictors across the two subject areas were sex and Black status. Hispanic status was common within the History assessment results, and ELP scores were common predictors of high-proficient EL response time in both subject areas. These results suggest that typically males have shorter processing time as compared to females, and Black students have longer processing time as compared to other ethnic groups. For Hispanic students, results were dependent on EL status, and as ELP score increased for high-proficient ELs, processing time decreased. These results agree with some of the response time literature where small differences in response time among subgroups have been found. Specifically, previous research has found that females take longer than males on assessments (Schaeffer et al., 1993), as do African American or Black students (O'Neill & Powers, 1993; Schaeffer et al., 1993), and Hispanic students (Llabre & Froman, 1987; O'Neill & Powers, 1993).

Understanding which predictors are likely to impact response time and ultimately processing time is essential to understand which factors could contribute to the differences in response time between ELs and non-ELs. These results are especially important since they identify which subgroups may be taking longer to process with test accommodations. These results are important to note when considering time limits on future assessments with and without test accommodations. Knowing that certain subgroups may take longer to process can help to reduce test scores being negatively affected for these subgroups by time restrictions.

5.3.3 Relationship between Student Performance and Response Time

Previous research on response time has suggested that high performing examinees tend to take longer on test items as compared to students with lower performance (Chang et al., 2005). Because of this finding, the relationship between student performance and response time across the original two test forms for the three student groups was investigated. History and Math results showed statistically significant moderate correlations between raw score and response time for non-ELs. Similarly, for both subject areas, statistically significant small correlations were found across the two EL groups. Additionally, statistically significant differences were found between non-EL and both

EL correlations on both forms for History, and for Form 1 on Math. On Math, only the difference between non-EL and mid-proficient EL correlations was statistically significant on Form 2. Additionally, a statistically significant small correlation was found for high-proficient ELs on Form 2, thus resulting in statistically significant differences in correlations between the two EL groups. These results indicate that as performance increased for student groups, response time also increased, consistent with previous research.

To further investigate this relationship, graphs were created showing individual item response time by raw score for all three student groups on accommodated and non-accommodated items. On the History assessment 14 items showed the non-EL response time increasing with increased test performance. For EL groups, 8 items showed some increase in response time with increased performance. Similar results were found on Math with around 15 items showing response time increasing with improved test performance for all three student groups. For the two EL groups on both subject areas, the response time trend across the raw score scale tended to be much more jagged as compared to the non-EL students. This was likely due to the smaller number of ELs at individual score points resulting in a less stable estimate of average response time. Similarly across all three student groups, there were small numbers of students at the extreme ends of the score scale, especially at the upper end for the EL student groups. Because of these small numbers, stable estimates of average response time on the extreme ends of the score scale were unattainable. Lack of response time estimates along the score scale could be part of the reason for not always seeing the changes in response time along the score scale.

Visually examining the item-level graphs allowed for inspection of different response times for the three different groups with accommodations and without. On most of the items across both subject areas, it was easy to see the non-EL group taking significantly less time to respond to items, and the similarities in response time between the two EL groups. Additionally, on some items it was possible to see that certain groups of students along the score scale took longer with the accommodated item as compared to the non-accommodated item, which was interesting. Most visually noticeable was the decreased response time on the last five items of the Math test, and the similarities in response time across the three student groups. The gap in response time between the EL groups and non-EL students declined. Although the tests were not specifically timed, it could be that students were still feeling some time pressure towards the end of the assessment and therefore rushed to complete the test items.

5.4 Accommodation Use Results in Relation to Response Time Results

It was speculated that response time increased as the number of accommodations available on item increased. For History, this trend was found for both EL groups, but not as much for non-ELs. Non-EL response time remained fairly stable regardless of the increase in number of accommodations. These results suggested that processing time did not change for ELs when accommodations were available, whereas, processing time increased for both EL groups as the number of accommodations increased. Thinking about this result in relation to the interaction hypothesis, it appears that the number of accommodations on a test item did not change the processing time for non-ELs, but did change the processing time for EL student groups. Hofstetter (2003) stated that an appropriate accommodation is one that produces this interaction effect, and even though this interaction effect is different than described in Hofstetter (2003), it is still an important interaction to note in relation to these test accommodations. On Math, however, no consistent trend was found for any of the student groups in relation to response time and the increased number of accommodations on an individual item, suggesting that processing time on individual items was not related to the number of accommodations available on that item. The result in Math suggests that student processing time might be more related to computation time rather than accommodation use, and that the difficulty of the test items may have played a larger role in processing time, hiding any potential affects on the accommodation.

To further investigate commonalities between accommodation use and response time, the statistically significant predictors on the accommodation use regression models were compared to the statistically significant predictors on the response time regression models. Table 5.1 shows the directions of the statistically significant predictors across the accommodation use, accommodated response time, and non-accommodated response time regression models for all three student groups across both subject areas. For non-ELs, sex and Hispanic status were statistically significant across the accommodation use and response time models. Sex was significant across both subjects, whereas Hispanic was only significant across History models. Sex and Hispanic were also significant predictors across all three models in History for high-proficient ELs. Additionally, SES was a statistically significant predictor across all three models in History for high-proficient ELs. For mid-proficient ELs, common statistically significant predictors across accommodation use and response time included Black and Native Hawaiian status for Math. As expected, there were more common predictors across accommodation use and response time in History than in Math. This was expected given that accommodations did not impact response time in Math, but did impact response time in History.

5.5 Validity of Computer-Based Test Accommodations

The *Standards for Educational and Psychological Testing* (1999) define validity as "the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests" (p. 9). In this study evidence based on response processes was gathered to evaluate the fit between the construct and responses actually engaged by the examinees (AERA et al., 1999). When ELs take an assessment with test accommodations, specifically direct linguistic accommodations, those test scores are interpreted in a way that suggests that the linguistic complexity has been reduced enough to provide accurate representations of EL student performance. Essentially, those test scores are interpreted in the same way as non-EL scores, assuming that the playing field for ELs has been leveled and that the construct-irrelevant variance for ELs has been removed with the accommodation provided. In providing evidence based on response processes, it is possible to identify if there are still potential sources of construct-irrelevant variance that could be impacting EL test scores.

This study shows how to use accommodation use and response time data to examine whether EL test scores are accurately representing "true" performance. On the History assessment, high percentages of ELs were accessing accommodations, and taking a longer time to process on many of the items with accommodations; however, because there were still high percentages of ELs not accessing the accommodations, the test scores of the EL group as a whole might not be an accurate representation of EL performance. This same result was noted for the Math assessment. Additionally, response time results in Math showed ELs taking the same amount of time to respond to items regardless of receiving a test accommodation. These results suggest either additional sources of construct-irrelevant variance not addressed through the accommodations, or ELs struggling with the test content.

At this juncture, results suggest that further investigation regarding the validity of the computer-based test accommodations in this study. However, for History, there was a notable relationship between number of accommodations on an item and response time. Specifically, non-EL response time did not change as accommodation numbers increased, whereas EL processing time did increase as the number of accommodations increased. This is a positive result for the History computer accommodations, suggesting that non-ELs are not processing test items any differently when receiving accommodations. For Math, however, it is likely that this trend was masked by the fact that the test items were difficult for students, regardless of the accommodation provided. In the next section, suggestions to continue with this research are provided.

5.6 Limitations and Directions for Future Research

Although results of this study can inform future research on computer-based test accommodation research, there are some limitations that should be discussed. Some important limitations are in relation to the specific data set used in this study. This data set only had two subject area assessments in History and Math, and each assessment only contained 25 test items. Additionally, there were only two types of accommodations available, and this study did not look at each accommodation specifically. Lastly, the sample involved mostly Hispanic ELs and only the high school grade level. That be said, the study was an experimental design, allowing for direct comparisons between ELs and non-ELs using and not using test accommodations, which was a major advantage to the study.

Given the limitations of this study, some suggestions for future research are presented. Specifically, future research should consider investigating tests of longer length, especially given the result that accommodation use decreased as the assessment progressed. This phenomenon would be better explained if the tests were of longer length. Because of the differences in results across History and Math in relation to accommodation use and response time, other test subjects and grade levels should be considered. Abedi et al. (2003b) found differences in accommodation across grade levels in relation to response time, so it would be interesting to see if that same result would appear if the study was replicated. Other accommodations should also be investigated, and investigated exclusively instead of together to more closely examine the validity and effectiveness of specific accommodations. Lastly, although a large population of ELs is Hispanic, accommodation research would benefit from considering other ethnic groups and languages spoken in the home, as well as breaking down the Hispanic students into different subpopulations (e.g., Mexican, Spanish, Puerto Rican, etc.).

The results of this study also bring forth suggestions for future research. One of the results of this study showed differences in accommodation use across the three student groups. Future research would benefit by looking more closely at the students actually using the test accommodations. The regression results suggested predictors of accommodation use, but it would be beneficial to look at the demographic make-up of students more generally. It would also be advantageous to break down student performance in relation to accommodation use. Did the students who actually used the accommodations perform higher than their peers who did not use the accommodations? Focusing on students using the accommodation, the interaction hypothesis and differential boost hypotheses could be examined to see if the results were different based on accommodation use. This breakdown would allow for a more accurate representation of accommodation effectiveness and validity.

This current study did not break down the results based on accommodations. Future research should replicate the accommodation use results by accommodation type and look at whether pop-up glossary accommodations were more often used than the sticker-tool accommodation. In relation to response time, future research would benefit by looking at how long students keep accommodations open. Although students are selecting accommodations, it could be that they accidentally clicked on the accommodation and immediately closed it. If the accommodation stayed open longer, it would suggest more processing time using the accommodation.

5.7 Conclusions

This study investigated the validity of computer-based test accommodations for ELs through evidence based on response processes. Accommodation use can inform whether students are actually accessing the accommodations. Results of this study showed around 36% of ELs not using accommodations on average for History and around 49% of ELs not using accommodations on average for Mathematics. Given that high percentages of ELs were not using the accommodations, it is difficult to see the true level of effectiveness. These results encourage future research to consider accommodation use when evaluating the effectiveness of test accommodations. Additionally, accommodation use declined as the assessment continued, which could have been a result of students not finding the accommodation useful at the beginning of the assessment. This could be problematic if tests are longer than 25 items and EL students are not getting the full benefit of the accommodation.

Response time results allowed for examination of student processing time with and without accommodations. Results of this study showed students processing longer with accommodations on the History assessment, but not on the Math assessment. These results could be due to the fact that History is more of a reading based assessment and therefore students were more likely to access or use the accommodations, whereas Math involved more computation. Additionally, the Math assessment could have simply been more difficult, resulting in similar response times regardless of receiving an accommodation.

This study tended to fill the void in EL computer-based test accommodation literature. EL accommodation literature has limited research available on computer-based accommodations, and has failed to provide additional sources of validity evidence, such as evidence based on response processes. This study showed ways to gather evidence based on response processes and described ways at which this information should be combined with current methods used in EL accommodation research. When students choose not to use accommodations that would help to reduce the language barrier, the scores they receive on content-area skills are likely to be less accurate. Additional research should look at finding ways to increase the relation of accommodation usage and score accuracy. If students are not using test accommodations available to them, the linguistic barrier is not being reduced, and test scores will not accurately reflect true student performance. As assessments shift to the computer, more information such as accommodation use and response time can be used to more appropriately examine the effectiveness and validity of test accommodations.

Table 5.1. Statistically Significant Predictors across All Regression Models

Predictors[a]	Non-ELs						High-Proficient ELs						Mid-Proficient ELs					
	Accom Use		RT: Accom		RT: Non-Accom		Accom Use		RT: Accom		RT: Non-Accom		Accom Use		RT: Accom		RT: Non-Accom	
	H	M	H	M	H	M	H	M	H	M	H	M	H	M	H	M	H	M
Sex	−	+	−	−	−	−	−		−		−				−	−		−
Hispanic	+		+		+		−		−		−	−		+				
Black			+	+	+	+				+				+	+	+	+	+
Asian			+		+													−
Am Indian[b]	−								−									
Native HI														+	+	+		+
SES							+	+	+		+							
At-Risk									−					+				
ELA Score[b]	−								−		−							
ELP Score										−	−	−						

Note. Accom = Accommodation; RT = Response Time; H = History; M = Math; Am = American; HI = Hawaiian; SES = Socioeconomic status; ELA = English Language Arts; ELP = English Language Proficiency.

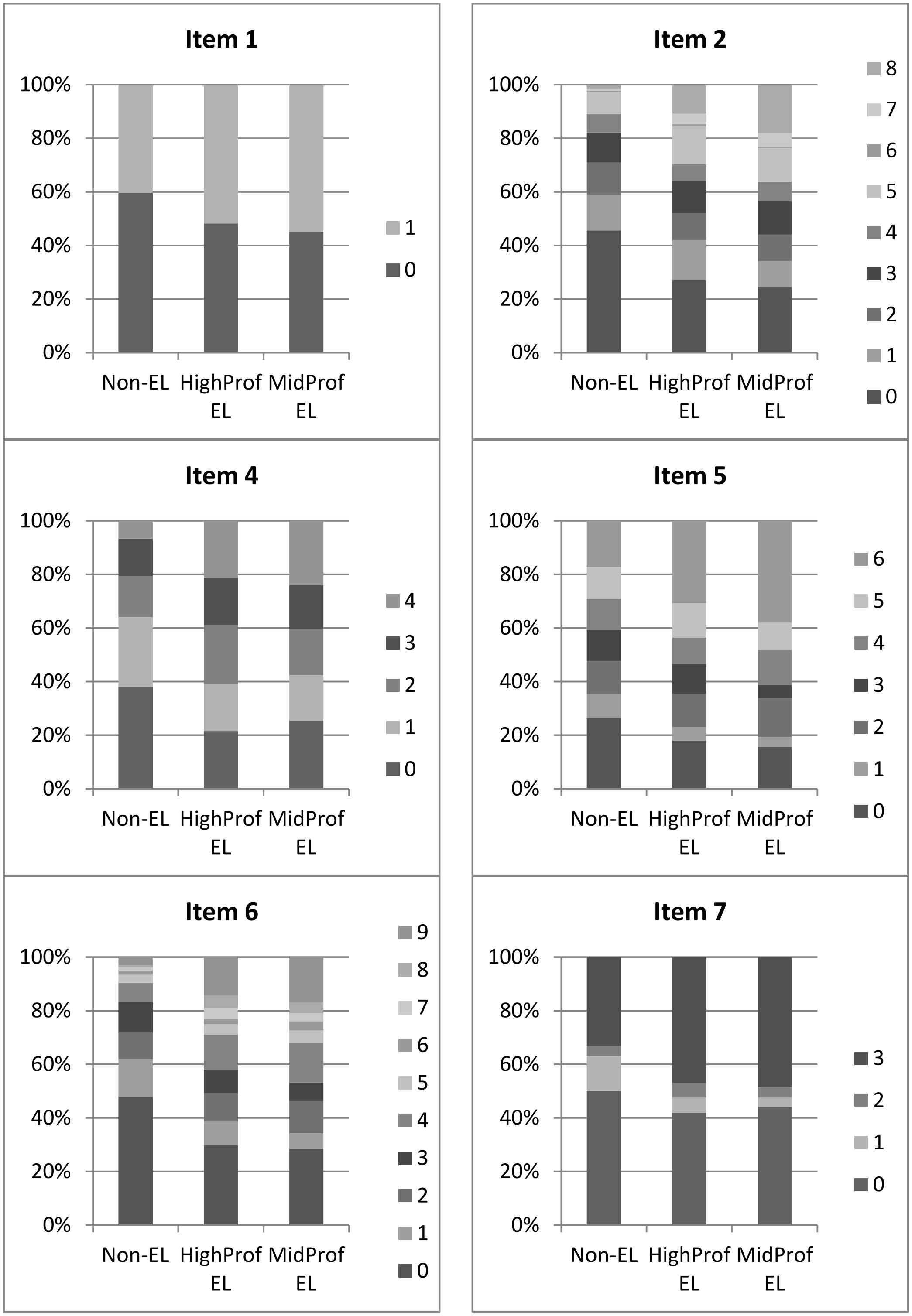
[a]Years was not a significant predictor in any of the regression models.

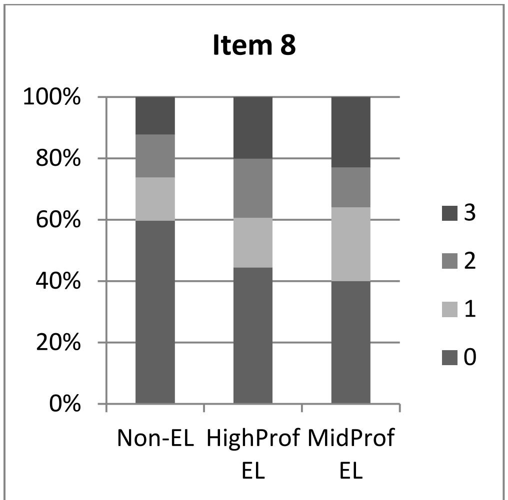
[b]Not significant predictors on any Math models.

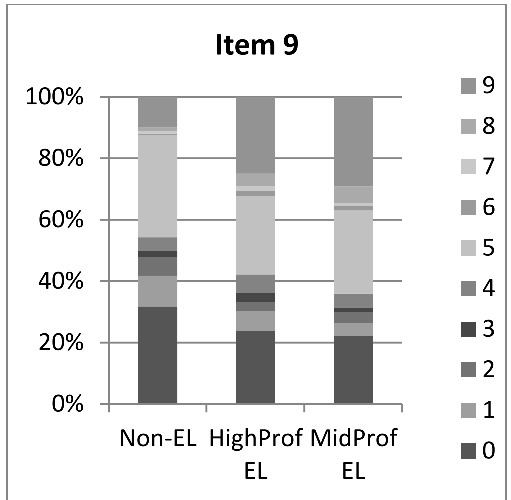
APPENDIX A

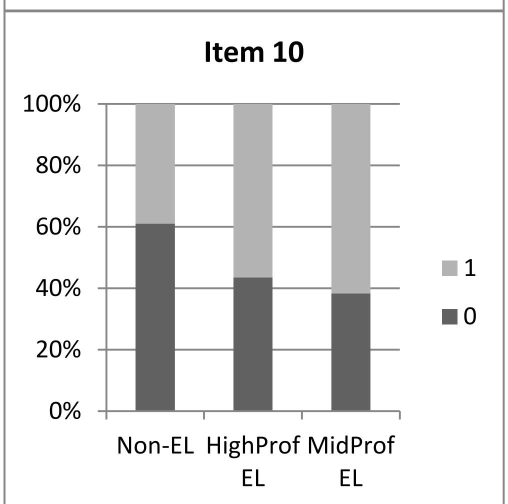
HISTORY ITEM-LEVEL ACCOMMODATION USE

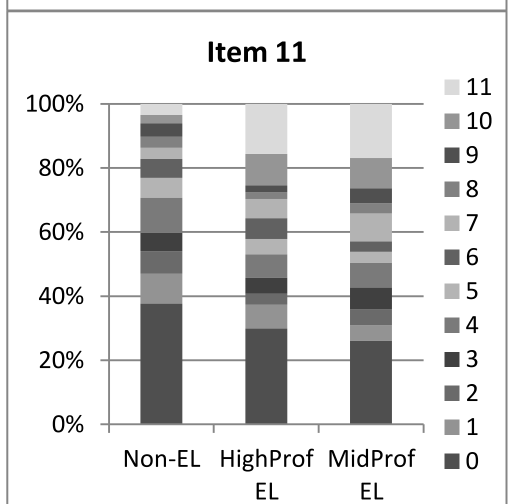
Table A.1. Percentage of Student Groups Using Different Numbers of Accommodations by Item

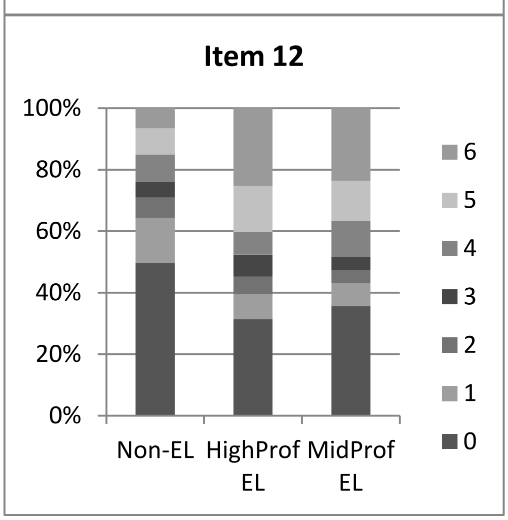


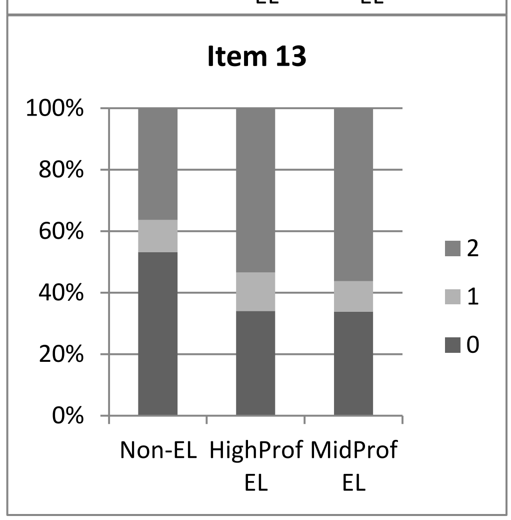

Item 8
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
3
2
1
0

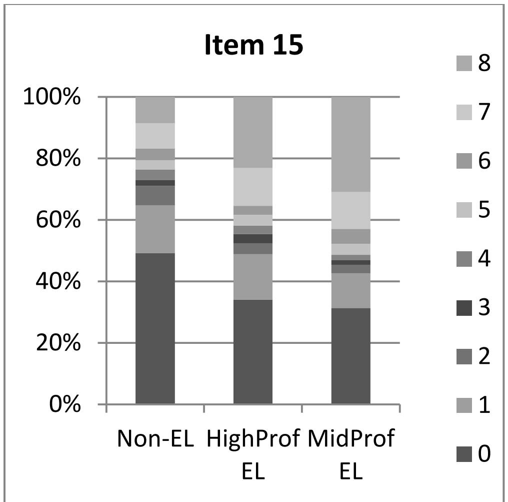

Item 9
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
9
8
7
6
5
4
3
2
1
0

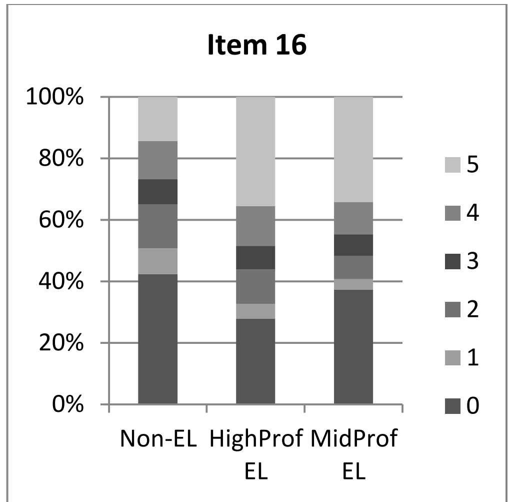

Item 10
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
1
0

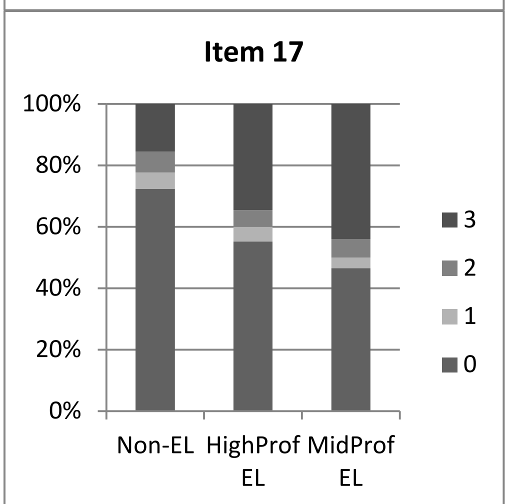

Item 11
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
11
10
9
8
7
6
5
4
3
2
1
0

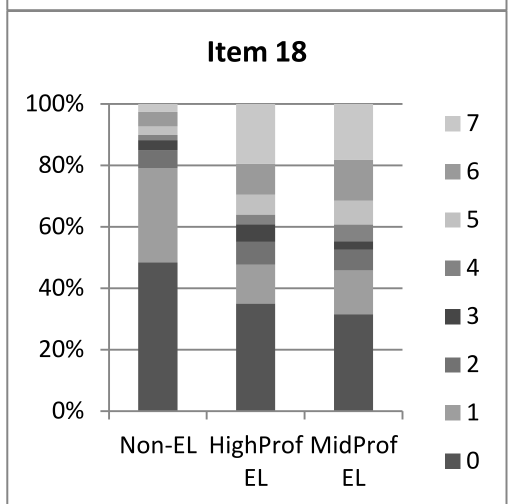

Item 12
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
6
5
4
3
2
1
0

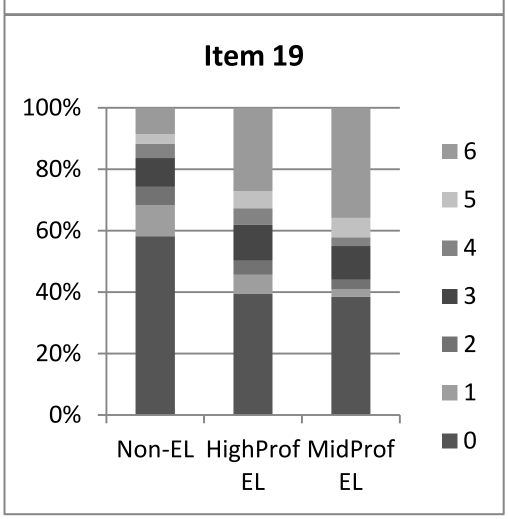

Item 13
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
2
1
0

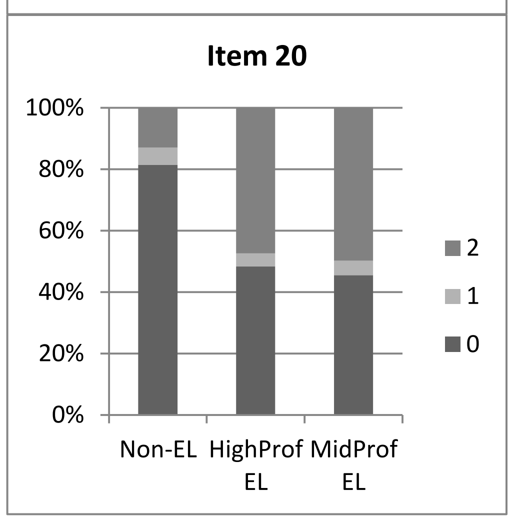

Item 15
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
8
7
6
5
4
3
2
1
0

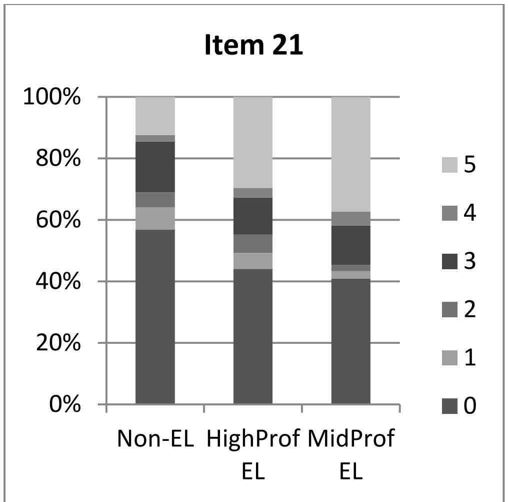

Item 16
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
5
4
3
2
1
0

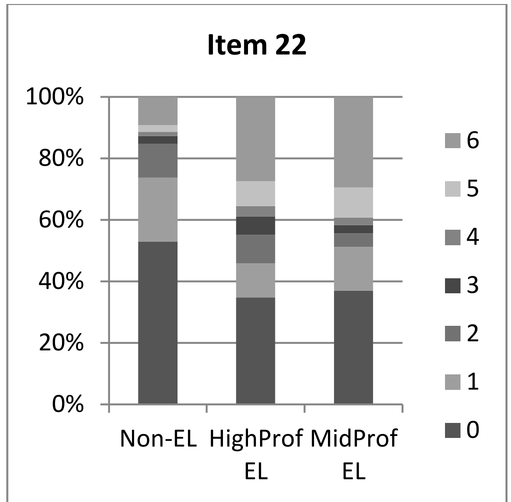

Item 17
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
3
2
1
0

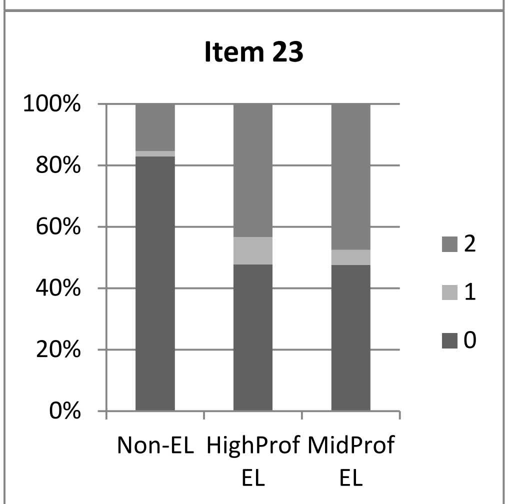

Item 18
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
7
6
5
4
3
2
1
0

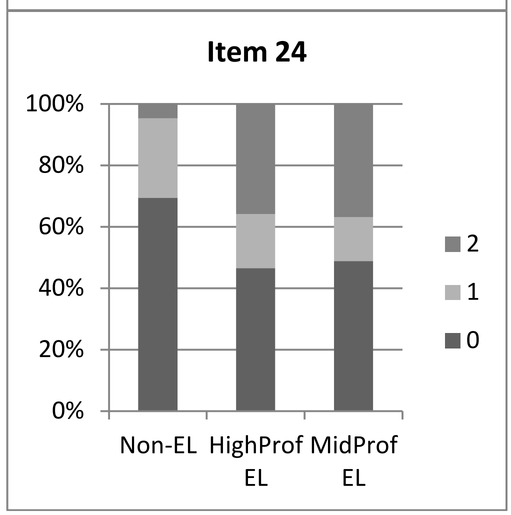

Item 19
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
6
5
4
3
2
1
0

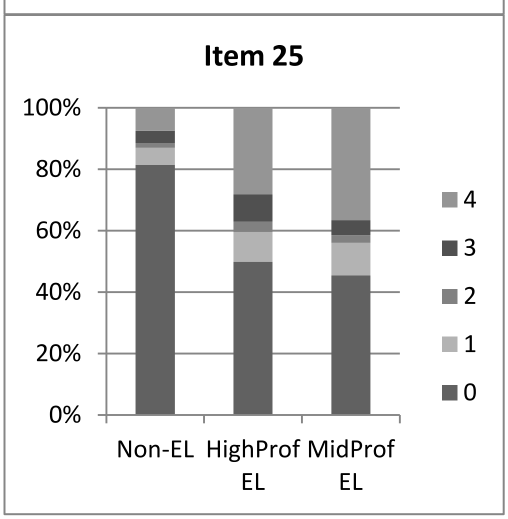

Item 20
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
2
1
0


Item 21
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
5
4
3
2
1
0


Item 22
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
6
5
4
3
2
1
0


Item 23
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
2
1
0


Item 24
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
2
1
0


Item 25
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
4
3
2
1
0

APPENDIX B

HISTORY ITEM-LEVEL RESPONSE TIME ACROSS FORMS 1A AND 2A

Table B.1. History Form 1A (Accommodated Form)

		Non-EL				High-Prof EL				Mid-Prof EL			
			RT (sec)	Exp(Ln(RT)) (sec)			RT (sec)	Exp(Ln(RT)) (sec)			RT (sec)	Exp(Ln(RT)) (sec)	
Item	# of Accoms	Prop Correct	Median	Mean	SD	Prop Correct	Median	Mean	SD	Prop Correct	Median	Mean	SD
1	1	.65	23.0	22.42	2.10	.66	32.0	34.12	2.23	.66	39.0	43.38	2.32
2	8	.71	29.0	33.12	2.16	.58	56.0	54.05	2.41	.51	67.0	64.07	2.39
4	4	.45	32.0	37.34	1.99	.50	57.0	54.60	2.51	.43	61.0	59.74	2.59
5	6	.64	40.0	53.52	2.34	.55	63.0	64.07	2.18	.34	76.0	73.70	2.14
6	9	.67	59.0	20.70	2.20	.42	107.0	95.58	2.56	.35	122.0	111.05	2.36
7	3	.57	20.0	21.54	1.95	.58	32.0	32.14	2.41	.51	37.0	36.97	2.29
8	3	.54	21.0	30.57	2.29	.55	39.0	38.47	2.27	.47	42.0	43.82	2.32
9	9	.73	34.0	21.12	2.08	.47	63.0	60.95	2.36	.31	85.5	77.48	2.39
10	1	.43	20.0	48.91	2.51	.26	39.0	42.10	2.48	.24	42.0	42.52	2.72
11	11	.76	58.0	36.97	2.29	.41	110.0	93.69	2.92	.34	120.0	108.85	2.56
12	6	.59	40.0	27.39	2.48	.42	76.0	73.70	2.51	.55	75.0	66.02	2.64
13	2	.47	30.0	33.12	2.16	.29	44.0	46.06	2.59	.23	50.0	50.40	2.75
15	8	.50	38.0	31.50	2.39	.34	71.0	65.37	2.64	.32	91.0	79.04	2.56
16	5	.48	44.0	36.23	2.48	.34	78.0	69.41	2.72	.29	72.0	64.72	3.10
17	3	.60	23.0	21.76	2.29	.44	39.0	38.86	2.66	.47	41.5	41.26	2.48
18	7	.62	31.0	28.22	2.41	.40	63.0	55.15	2.92	.40	69.0	56.26	3.03
19	6	.34	23.0	21.12	2.53	.24	45.0	42.10	2.86	.23	51.5	46.06	2.94
20	2	.40	27.0	24.78	2.29	.29	48.0	48.91	2.23	.26	47.0	49.40	2.34
21	5	.39	25.0	21.98	2.46	.28	43.5	39.65	2.92	.31	54.5	47.94	2.64
22	6	.37	26.0	23.10	2.56	.24	45.0	40.04	3.00	.24	45.0	40.04	3.03
23	2	.68	14.0	13.74	2.01	.59	23.0	24.53	2.51	.46	25.0	26.05	2.32
24	2	.74	13.0	13.07	2.03	.43	24.0	25.53	2.56	.30	29.0	29.67	2.86
25	4	.81	13.0	13.20	2.12	.61	27.0	28.50	2.59	.54	34.0	34.81	2.75

Note. Prop = Proportion.

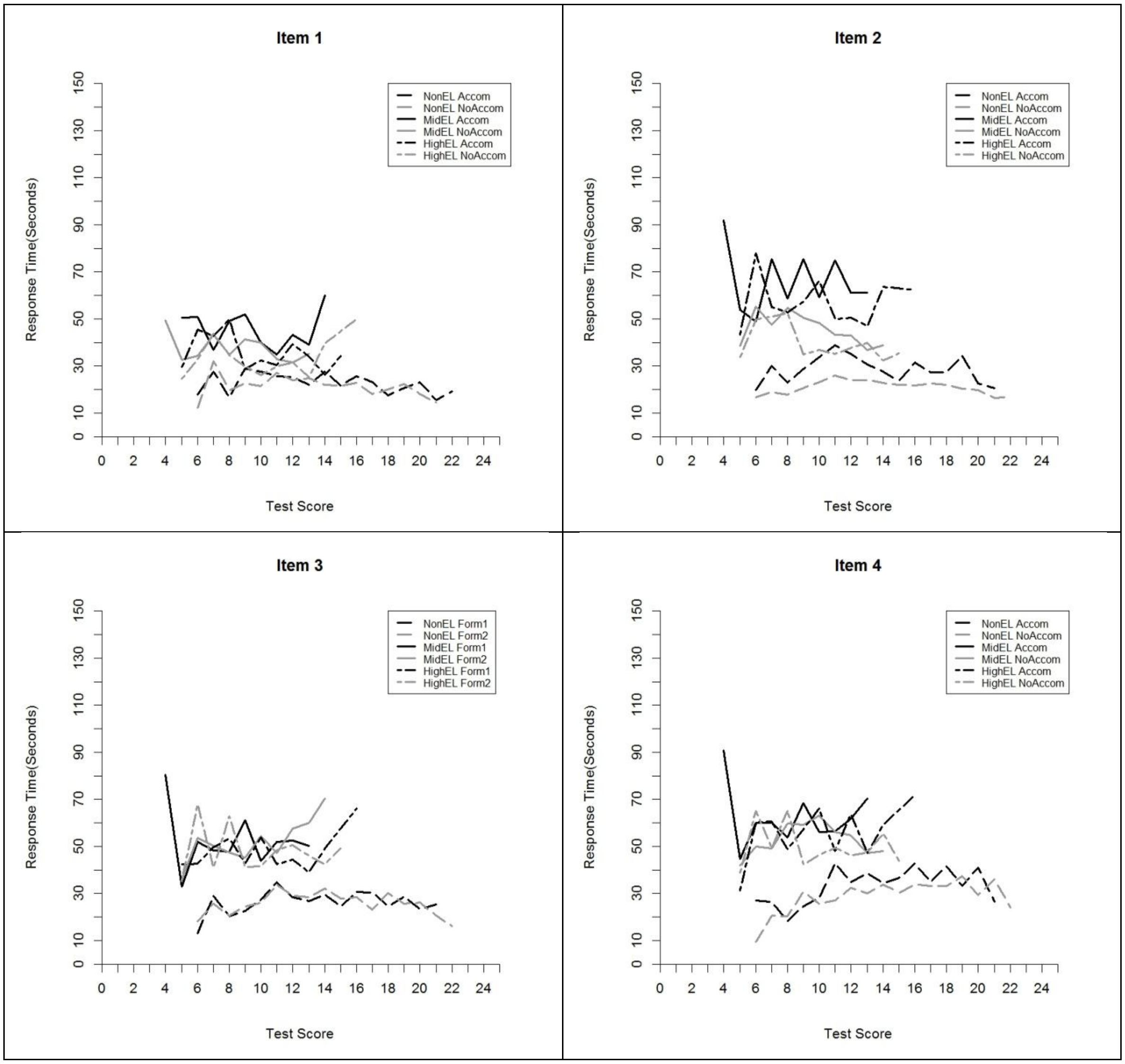
Table B.2. History Form 2A (Non-Accommodated Form)

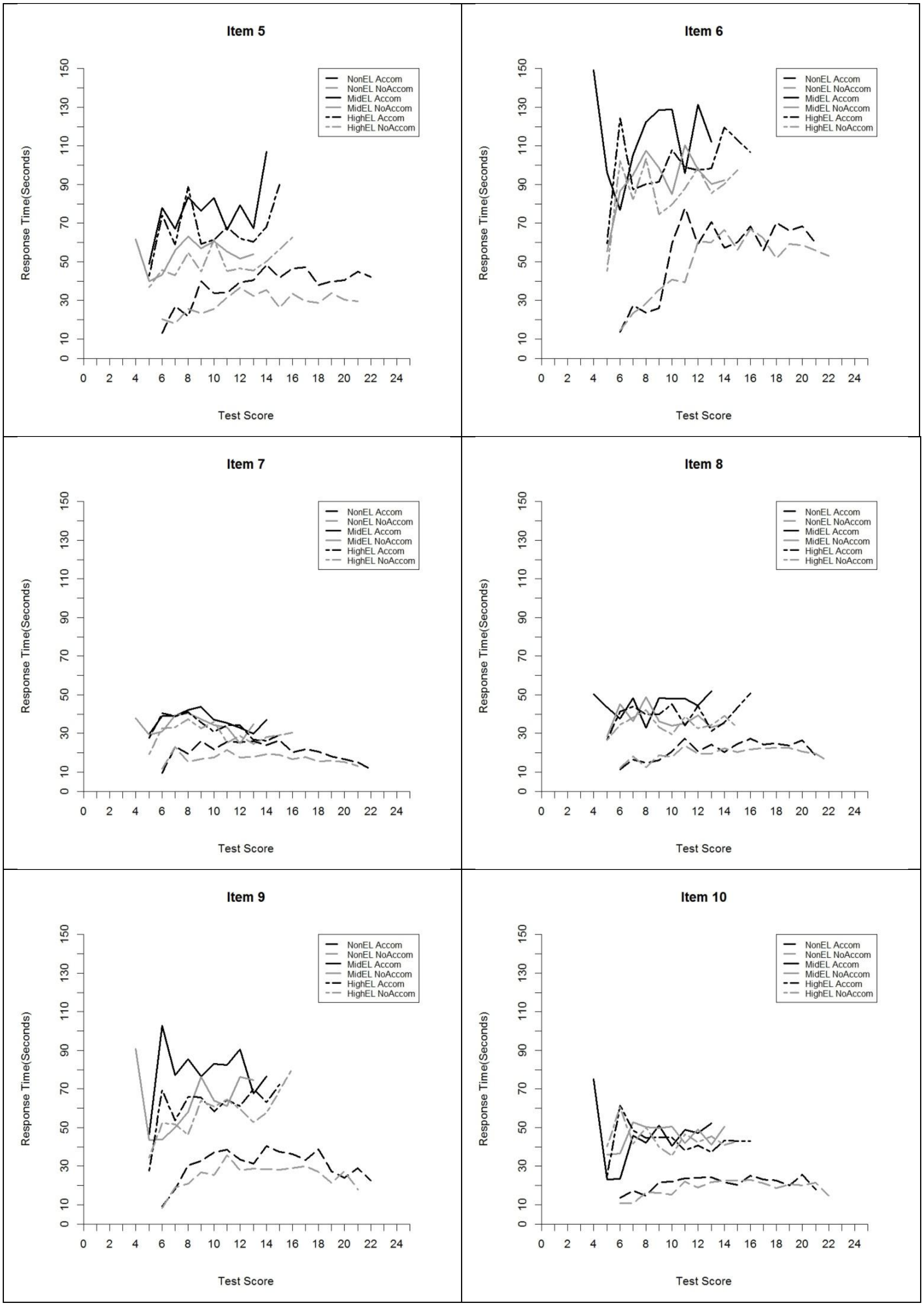
	Non-EL				High-Prof EL				Mid-Prof EL			
	Prop	RT (sec)	Exp(Ln(RT)) (sec)		Prop	RT (sec)	Exp(Ln(RT)) (sec)		Prop	RT (sec)	Exp(Ln(RT)) (sec)	
Item	Correct	Median	Mean	SD	Correct	Median	Mean	SD	Correct	Median	Mean	SD
1	.70	21.0	21.33	2.16	.69	31.0	31.50	2.44	.61	36.0	36.60	2.34
2	.69	20.0	20.91	2.05	.61	37.0	38.09	2.32	.64	44.0	46.06	2.25
3	.78	27.0	26.05	2.01	.46	44.0	45.60	2.29	.38	49.0	50.40	2.39
4	.42	28.0	29.08	2.25	.41	47.0	47.47	2.46	.40	55.0	54.60	2.39
5	.52	28.0	28.79	2.01	.35	45.0	48.42	2.39	.27	58.0	54.60	2.53
6	.62	56.0	46.53	2.66	.44	92.0	84.77	2.56	.35	98.0	93.69	2.46
7	.58	16.0	17.12	2.03	.57	28.0	29.67	2.46	.51	29.0	33.78	2.59
8	.51	18.0	19.49	2.12	.56	31.0	33.78	2.32	.49	34.0	36.97	2.16
9	.62	24.0	24.78	2.23	.26	53.0	55.15	2.83	.25	63.0	60.34	2.66
10	.46	19.0	18.73	2.25	.27	39.0	41.26	2.59	.22	46.5	46.99	2.72
11	.70	46.0	40.85	2.41	.40	83.0	75.94	2.89	.28	93.0	79.04	3.25
12	.64	35.0	32.46	2.39	.49	58.0	57.97	2.56	.62	63.5	63.43	2.56
13	.45	17.0	27.94	2.32	.24	28.5	47.47	2.66	.19	36.0	49.90	2.72
14	.42	28.0	17.64	2.10	.19	48.0	31.50	2.48	.22	52.0	36.60	2.56
15	.53	32.0	29.08	2.20	.38	60.0	55.70	2.69	.32	63.0	57.40	2.89
16	.42	32.0	25.28	2.69	.31	64.0	55.15	3.13	.30	69.0	59.74	3.06
17	.56	23.0	22.87	2.29	.47	36.0	37.34	2.59	.49	35.0	38.86	3.00
18	.59	24.0	22.65	2.59	.34	47.0	42.95	3.10	.36	48.0	40.85	2.89
19	.37	20.0	20.29	2.48	.28	36.0	36.97	2.86	.24	40.0	40.04	3.32
20	.32	24.0	22.42	2.36	.26	44.0	42.95	2.25	.23	46.0	44.26	2.23
21	.47	21.0	20.70	2.39	.29	39.0	39.25	2.94	.29	36.0	36.23	3.06
22	.39	21.0	18.73	2.51	.22	37.0	36.60	3.25	.25	38.5	35.52	3.06
23	.65	12.0	12.68	2.08	.55	23.0	25.03	2.56	.49	23.0	24.53	2.66
24	.71	11.0	11.25	1.99	.39	20.5	23.34	2.83	.27	22.0	23.81	2.61
25	.78	12.0	12.18	2.01	.64	23.0	25.79	2.69	.49	27.0	29.96	2.86

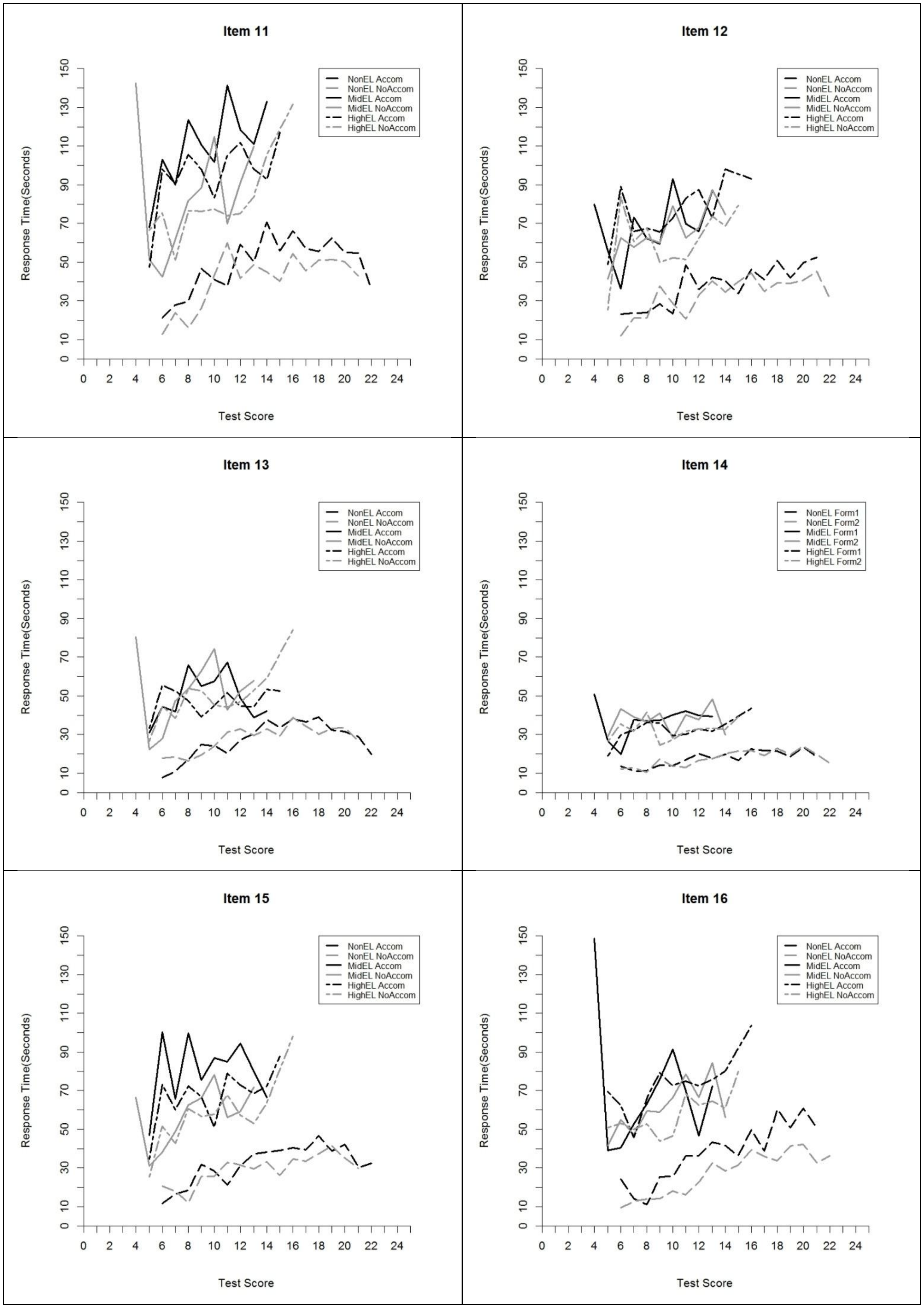
Note. Prop = Proportion.

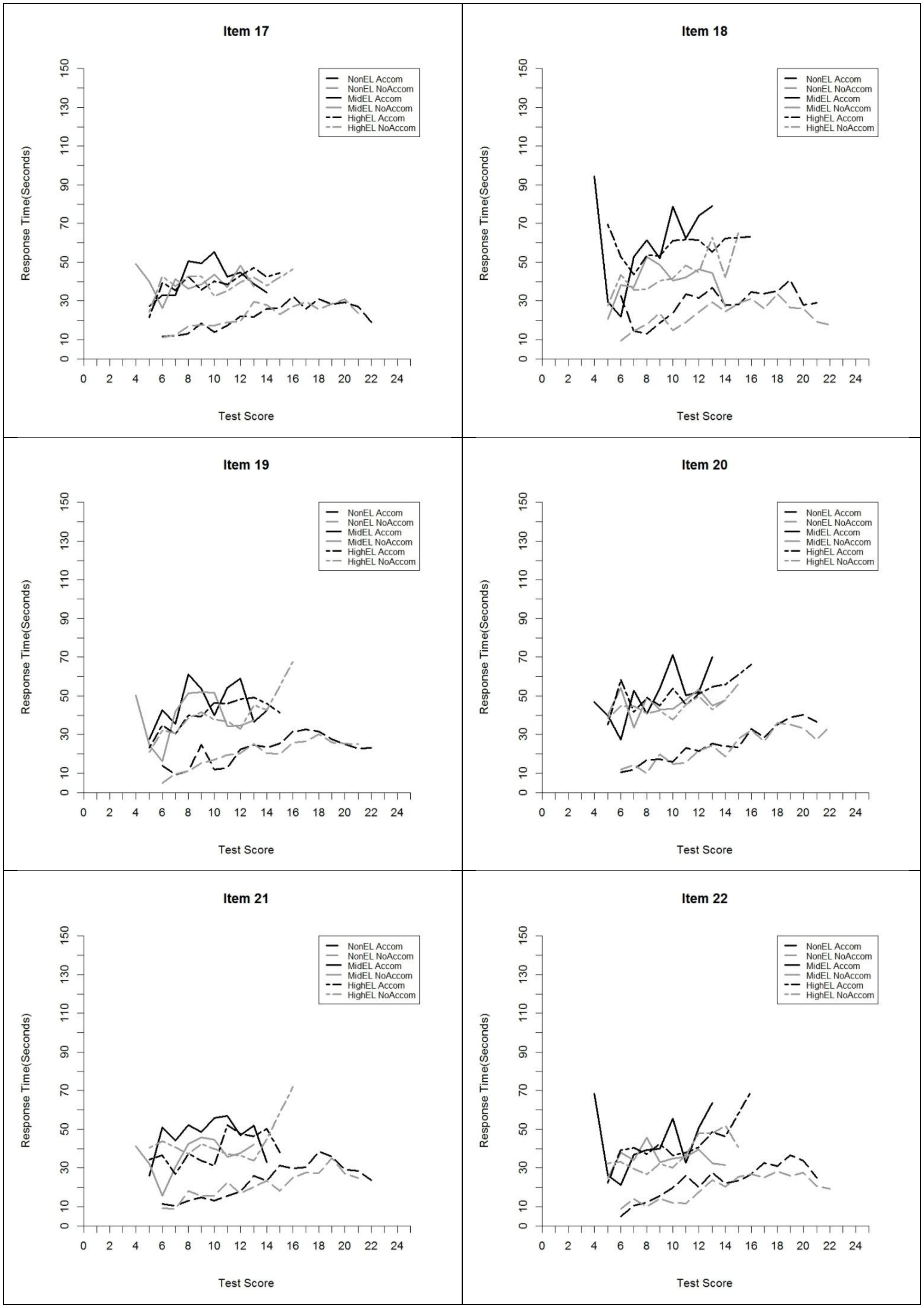
APPENDIX C

HISTORY ITEM-LEVEL RESPONSE TIME (SECONDS) GRAPHS

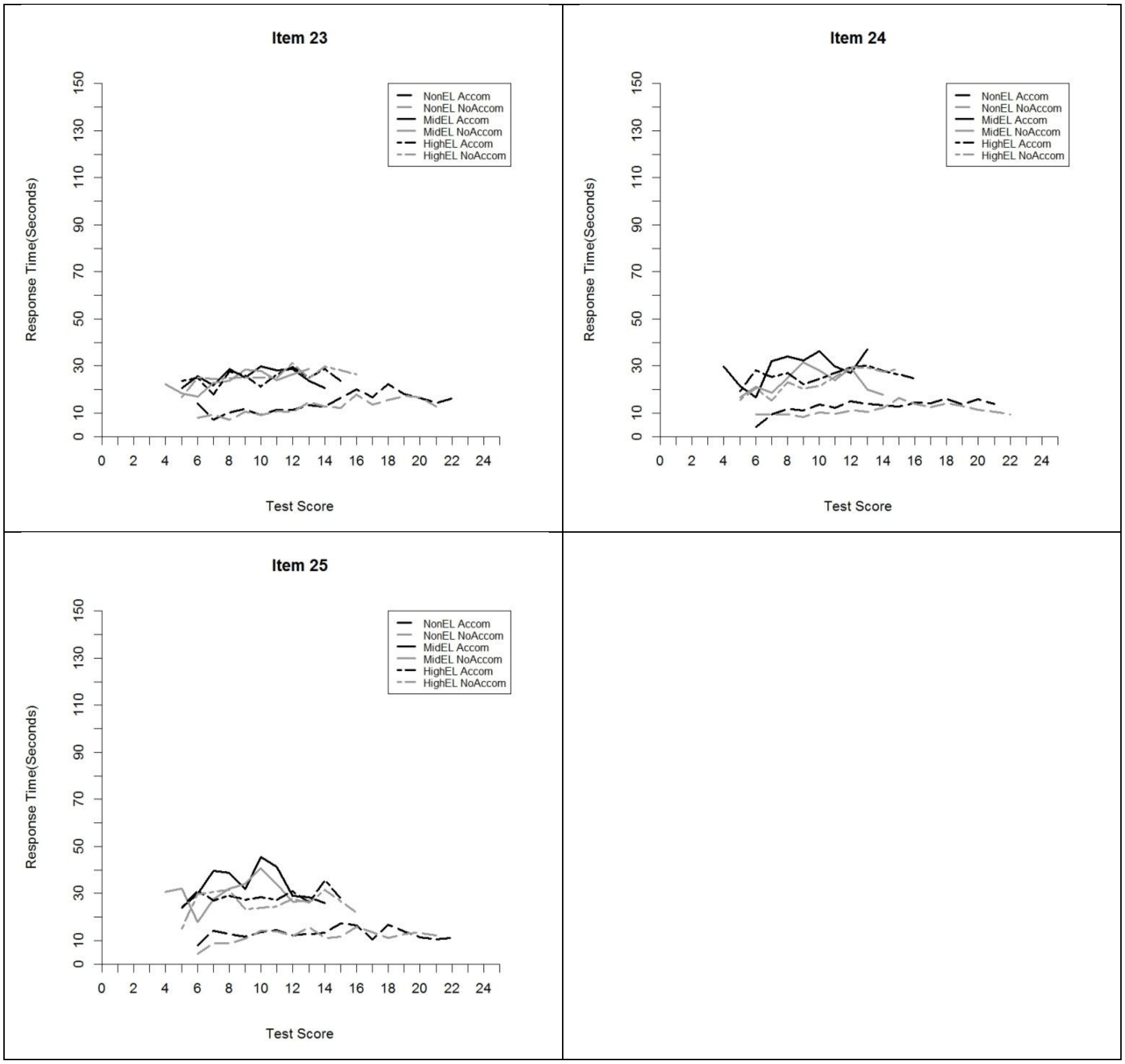



Item 5
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Item 6
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Item 7
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Item 8
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Item 9
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Item 10
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom


Item 11
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Item 12
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Item 13
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Item 14
Response Time(Seconds)
Test Score
NonEL Form1
NonEL Form2
MidEL Form1
MidEL Form2
HighEL Form1
HighEL Form2
Item 15
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Item 16
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom



Item 17
Item 18
Item 19
Item 20
Item 21
Item 22
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom

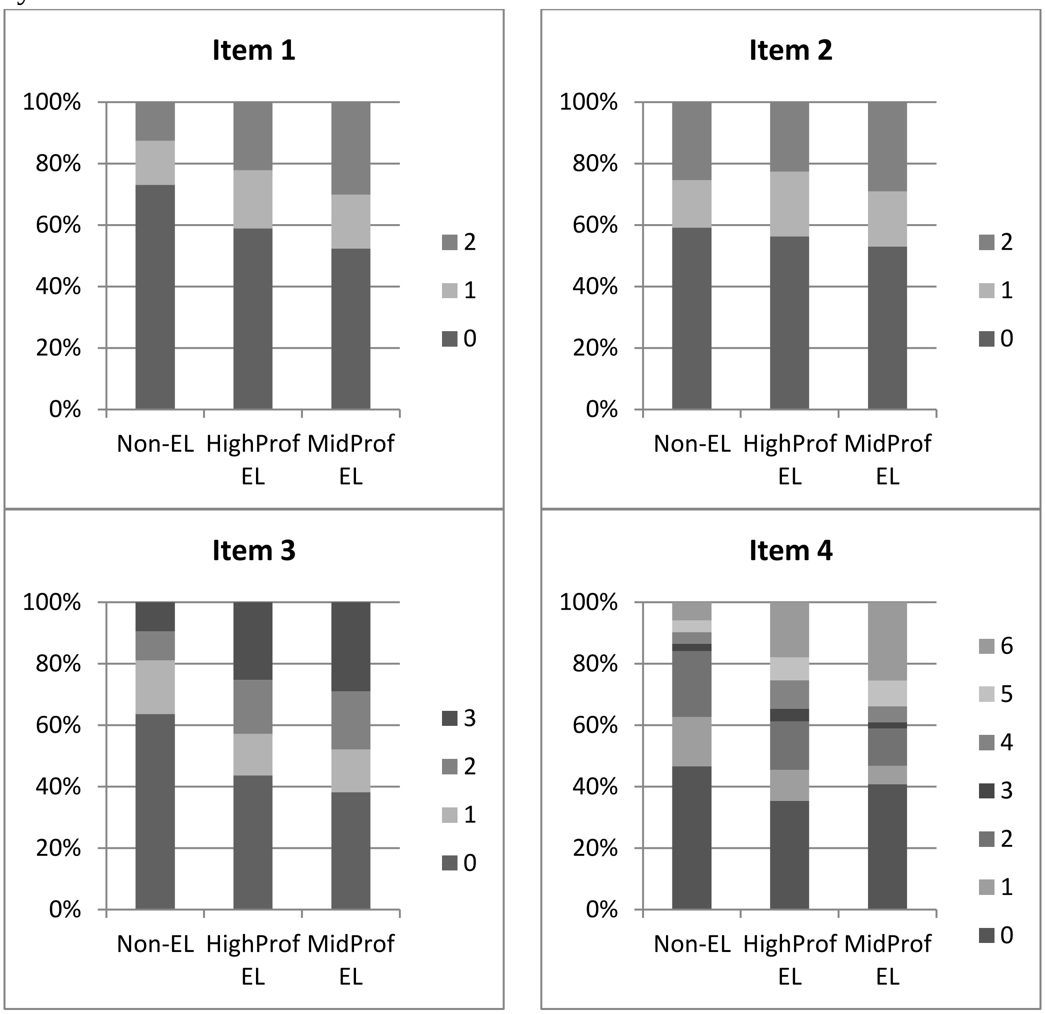


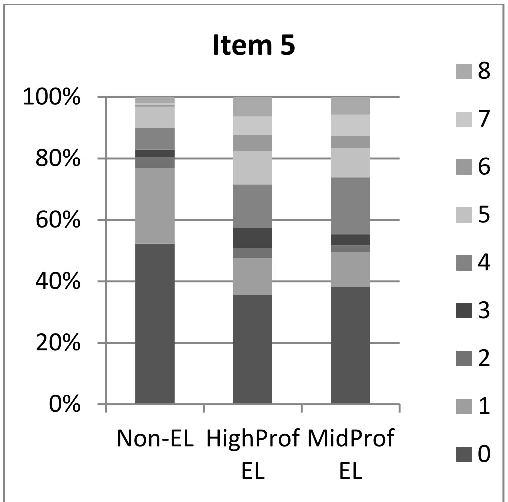
Item 23
Item 24
Item 25
Response Time(Seconds)
Test Score
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom

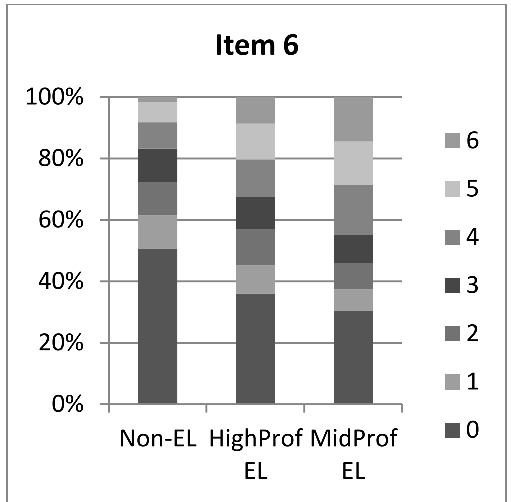
APPENDIX D

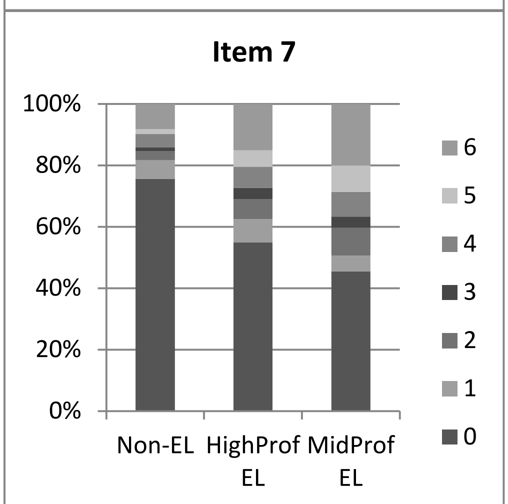
MATHEMATICS ITEM-LEVEL ACCOMMODATION USE

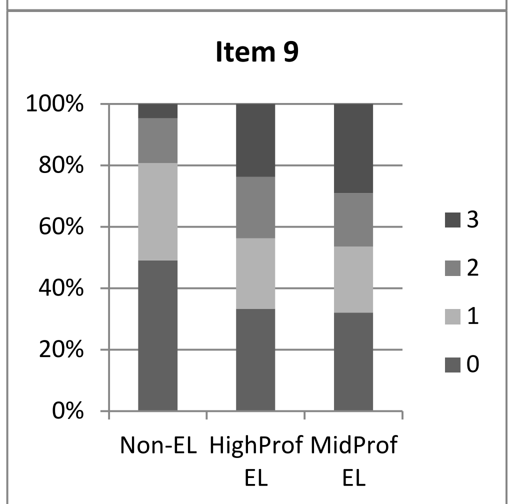
Table D.1. Percentage of Student Groups Using Different Numbers of Accommodations by Item

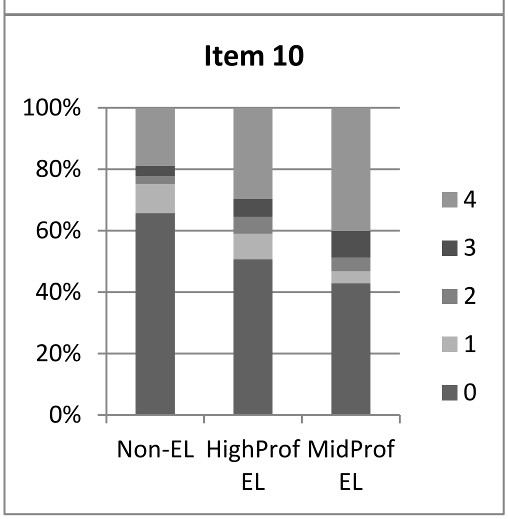


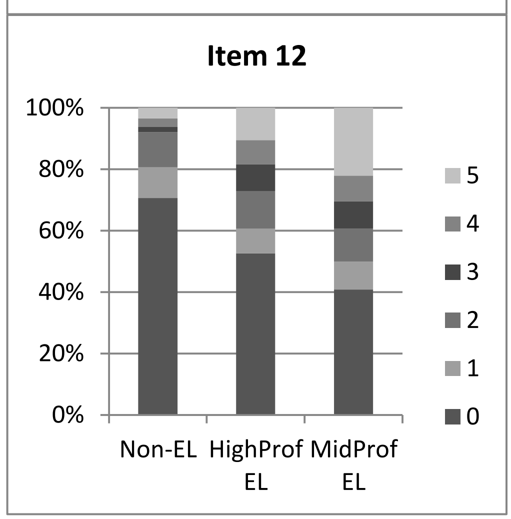

Item 5
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
8
7
6
5
4
3
2
1
0

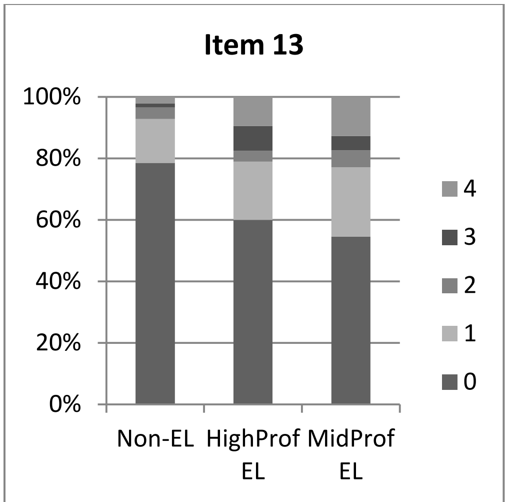

Item 6
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
6
5
4
3
2
1
0

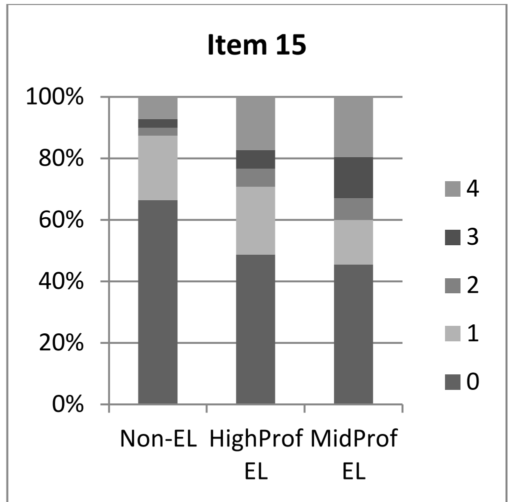

Item 7
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
6
5
4
3
2
1
0

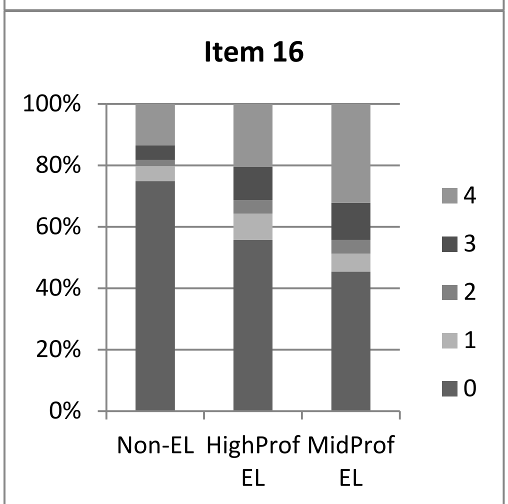

Item 9
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
3
2
1
0

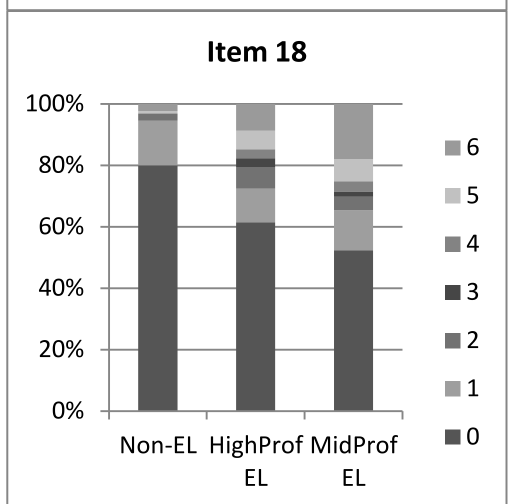

Item 10
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
4
3
2
1
0

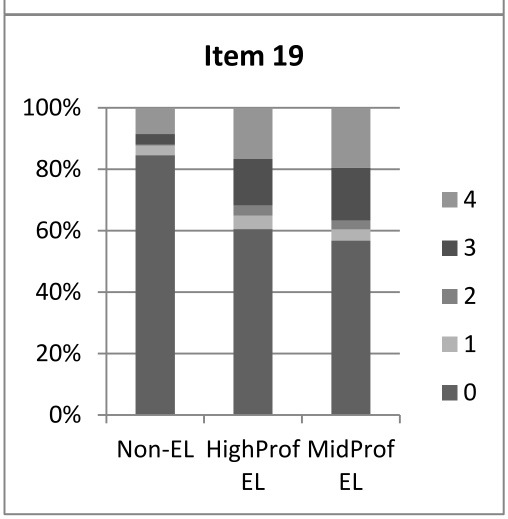

Item 12
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
5
4
3
2
1
0

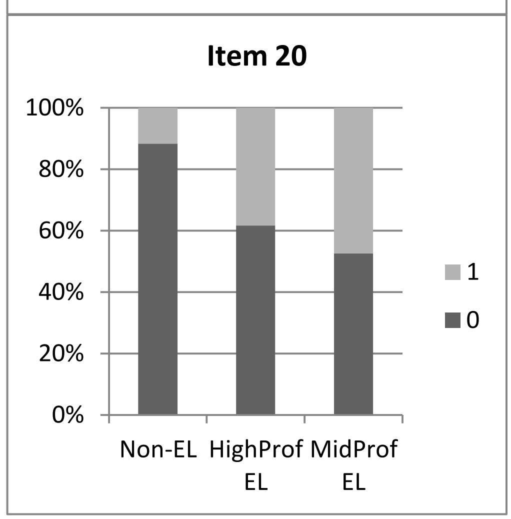

Item 13
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
4
3
2
1
0

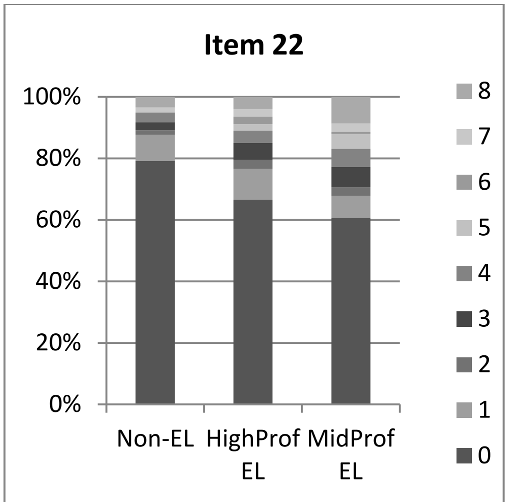

Item 15
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
4
3
2
1
0

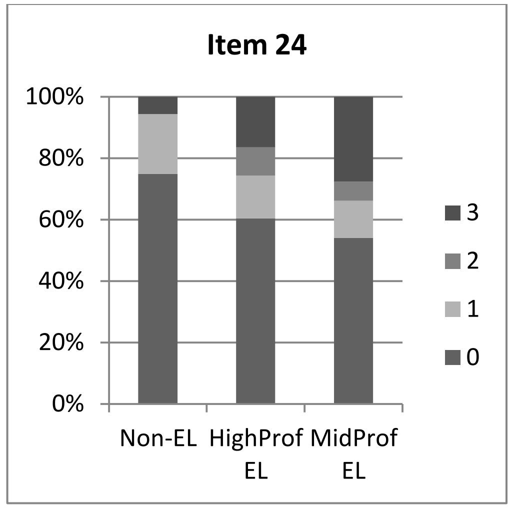

Item 16
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
4
3
2
1
0

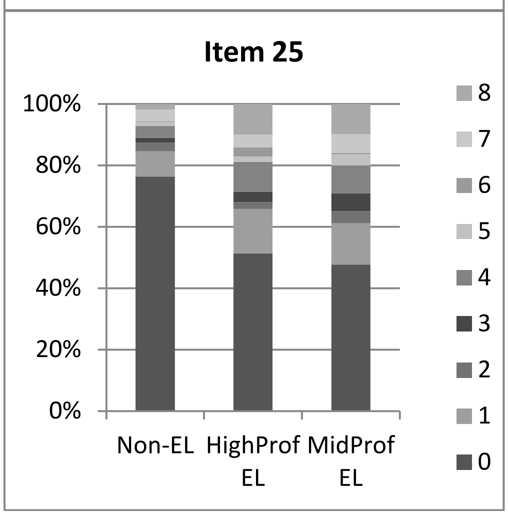

Item 18
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
6
5
4
3
2
1
0


Item 19
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
4
3
2
1
0


Item 20
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
1
0


Item 22
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
8
7
6
5
4
3
2
1
0


Item 24
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
3
2
1
0


Item 25
100%
80%
60%
40%
20%
0%
Non-EL
HighProf EL
MidProf EL
8
7
6
5
4
3
2
1
0

APPENDIX E

MATHEMATICS ITEM-LEVEL RESPONSE TIME ACROSS FORMS 1A AND 2A

Table E.1. Math Form 1A (Accommodated Form)

		Non-EL				High-Prof EL				Mid-Prof EL			
			RT (sec)	Exp(Ln(RT)) (sec)			RT (sec)	Exp(Ln(RT)) (sec)			RT (sec)	Exp(Ln(RT)) (sec)	
Item	# of Accoms	Prop Correct	Median	Mean	SD	Prop Correct	Median	Mean	SD	Prop Correct	Median	Mean	SD
1	2	.79	74.0	71.80	1.83	.45	129.0	123.95	2.10	.35	143.0	139.48	2.01
2	2	.46	108.0	99.15	2.58	.42	176.0	156.23	3.08	.37	198.0	183.11	2.41
3	3	.36	74.5	64.07	2.49	.20	110.0	101.04	2.64	.16	116.0	99.38	2.62
4	6	.56	148.0	119.38	3.14	.48	239.0	188.48	3.39	.40	220.0	190.15	2.94
5	8	.73	47.0	42.94	2.17	.36	97.0	88.11	2.62	.34	110.0	94.97	2.77
6	6	.45	66.5	57.48	2.46	.31	105.0	89.75	3.10	.33	119.0	106.59	2.48
7	6	.30	110.0	96.99	2.91	.24	190.0	160.98	2.97	.26	183.0	160.66	2.90
9	3	.68	63.0	56.89	2.69	.47	89.0	79.32	2.84	.42	88.0	79.04	2.69
10	4	.42	68.5	60.90	2.50	.36	114.5	92.93	3.20	.30	128.0	102.36	3.06
12	5	.28	78.0	58.70	3.10	.22	120.5	96.83	3.35	.15	124.0	108.95	2.79
13	4	.47	104.0	69.58	3.82	.28	175.0	141.09	3.59	.25	162.5	128.92	3.35
15	4	.29	58.0	50.64	2.84	.25	106.0	93.37	2.68	.22	99.5	87.83	2.80
16	4	.37	60.0	51.08	3.07	.26	111.5	87.14	3.71	.19	112.0	92.39	3.36
18	6	.31	46.0	37.89	3.44	.15	85.0	61.72	3.83	.13	83.0	69.16	3.40
19	4	.39	55.0	40.61	3.77	.26	90.0	77.66	3.57	.25	82.5	70.49	3.58
20	1	.29	29.0	25.03	2.37	.30	44.0	40.96	2.74	.25	50.0	49.26	2.81
22	8	.77	39.0	29.14	3.30	.34	68.5	54.30	3.58	.22	84.0	62.72	3.46
24	3	.60	16.5	15.82	2.14	.52	35.0	31.23	3.02	.38	36.0	32.48	3.16
25	8	.42	29.0	26.48	2.59	.19	56.0	54.03	2.81	.12	57.0	55.78	3.07

Note. Prop = Proportion.

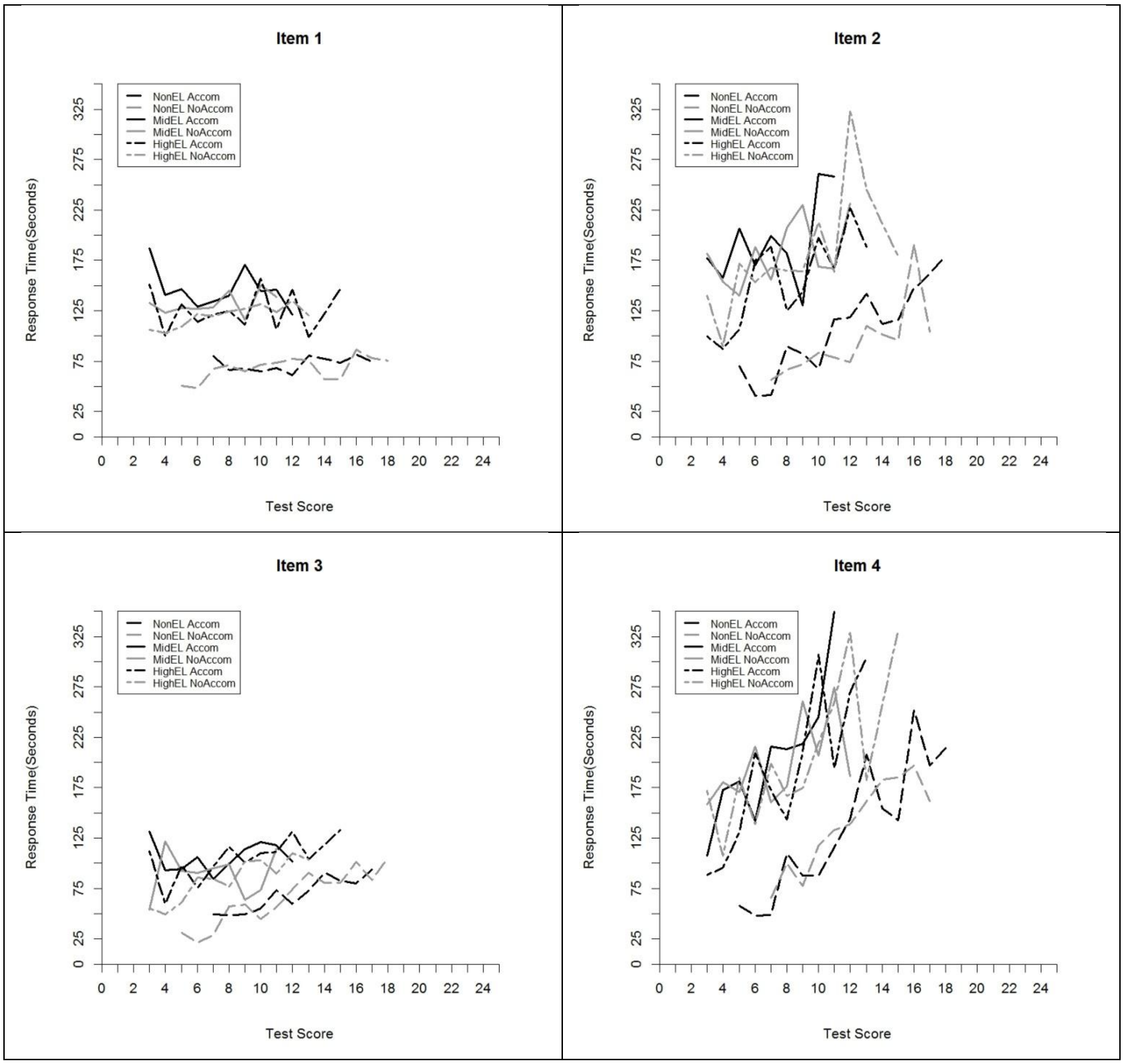
Table E.2. Math Form 2A (Non-Accommodated Form)

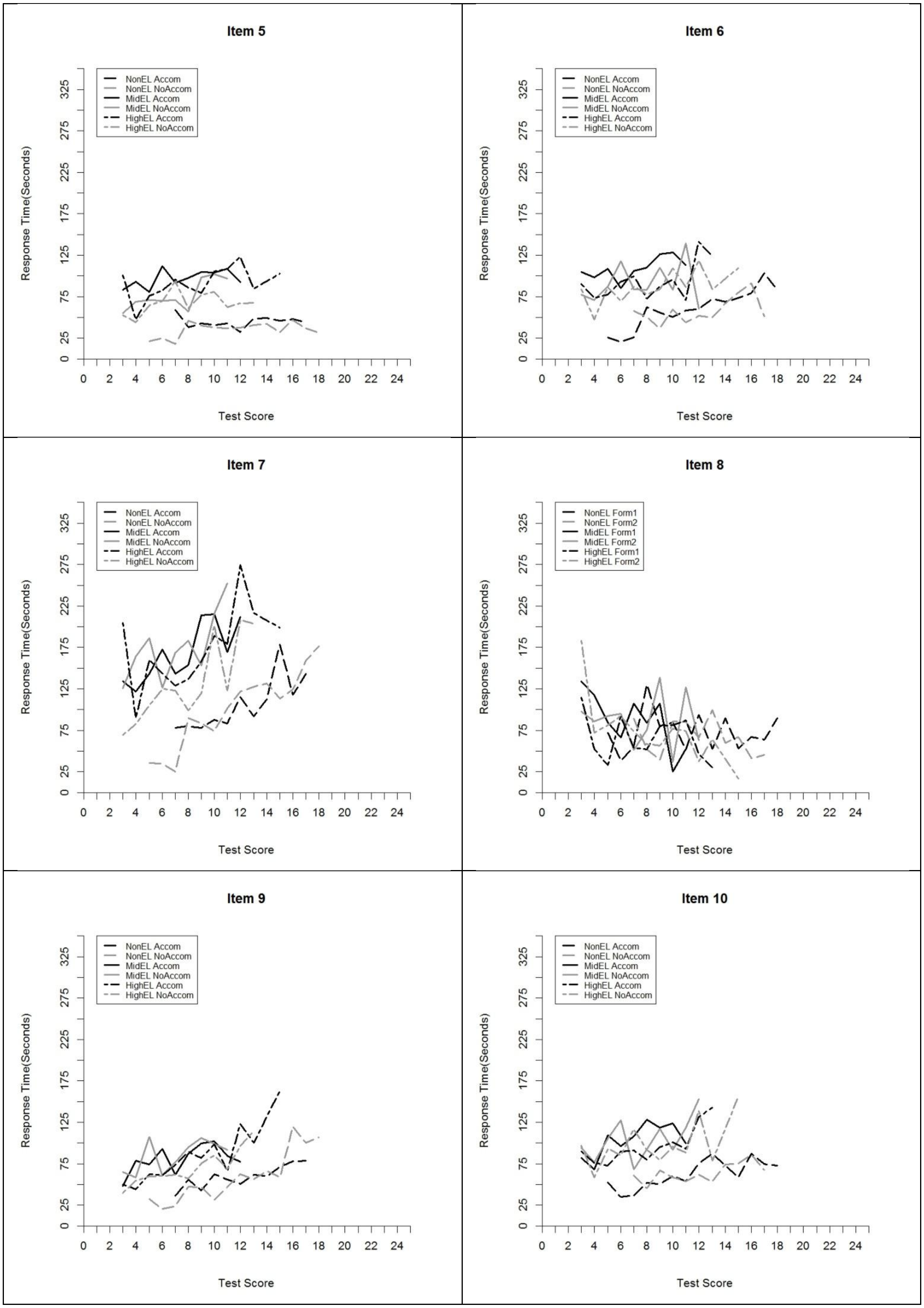
	Non-EL				High-Prof EL				Mid-Prof EL			
		RT (sec)	Exp(Ln(RT)) (sec)			RT (sec)	Exp(Ln(RT)) (sec)			RT (sec)	Exp(Ln(RT)) (sec)	
Item	Prop Correct	Median	Mean	SD	Prop Correct	Median	Mean	SD	Prop Correct	Median	Mean	SD
1	.78	73.0	69.25	1.92	.42	127.0	121.74	2.27	.31	132.0	126.71	1.99
2	.46	96.5	86.85	2.81	.49	194.0	174.64	2.67	.38	201.0	172.25	2.74
3	.33	66.0	60.50	2.63	.21	96.0	85.97	2.89	.15	98.0	88.95	2.60
4	.55	154.0	128.24	3.00	.45	220.0	189.05	2.98	.39	215.0	188.50	2.92
5	.66	38.0	35.97	2.20	.41	78.0	70.59	2.86	.33	81.0	72.56	2.67
6	.48	58.0	55.38	2.53	.28	92.0	84.31	2.75	.23	97.0	89.30	2.67
7	.23	108.5	91.43	3.09	.26	170.0	128.74	3.66	.23	200.0	165.42	2.82
8	.19	110.0	71.16	4.61	.10	122.0	62.71	7.32	.14	147.0	79.18	6.87
9	.65	56.0	51.69	3.15	.45	87.5	70.00	3.19	.39	91.0	79.82	2.94
10	.42	64.5	59.01	2.36	.36	113.0	93.95	2.86	.30	116.0	95.79	2.70
11	.44	61.0	49.70	3.20	.42	101.0	82.67	3.47	.32	96.0	80.77	3.34
12	.30	80.0	64.11	2.91	.25	117.0	102.90	2.80	.16	106.5	95.47	3.02
13	.45	18.0	65.13	3.82	.24	20.0	122.69	4.11	.23	24.0	107.07	4.02
14	.47	97.0	19.65	3.30	.40	178.5	20.59	4.20	.36	154.0	24.24	4.13
15	.31	50.0	45.12	3.01	.25	92.0	76.81	3.09	.24	95.0	76.50	3.00
16	.34	55.0	45.98	3.18	.30	101.0	88.70	3.41	.22	110.0	94.70	3.32
17	.20	27.0	27.58	2.92	.22	55.0	50.40	3.20	.16	53.0	48.38	2.94
18	.33	43.0	36.18	3.15	.12	79.0	70.64	3.22	.11	83.5	69.31	3.57
19	.37	45.0	32.05	3.69	.27	78.0	58.69	4.21	.26	78.0	59.16	3.89
20	.28	27.0	25.72	2.31	.32	47.0	45.29	2.35	.26	47.5	48.56	2.59
21	.53	25.0	19.48	3.17	.40	44.0	37.40	3.40	.38	43.0	38.92	3.35
22	.79	35.0	27.55	3.30	.26	75.0	63.80	3.19	.20	73.0	64.67	3.19
23	.82	20.0	19.39	2.35	.79	29.0	31.41	2.76	.76	28.0	29.24	2.79
24	.60	16.0	15.21	2.10	.47	34.0	34.72	2.91	.41	35.0	34.72	2.95
25	.36	24.0	23.58	2.54	.17	46.5	43.12	3.30	.11	47.0	46.04	2.98

Note. Prop = Proportion.

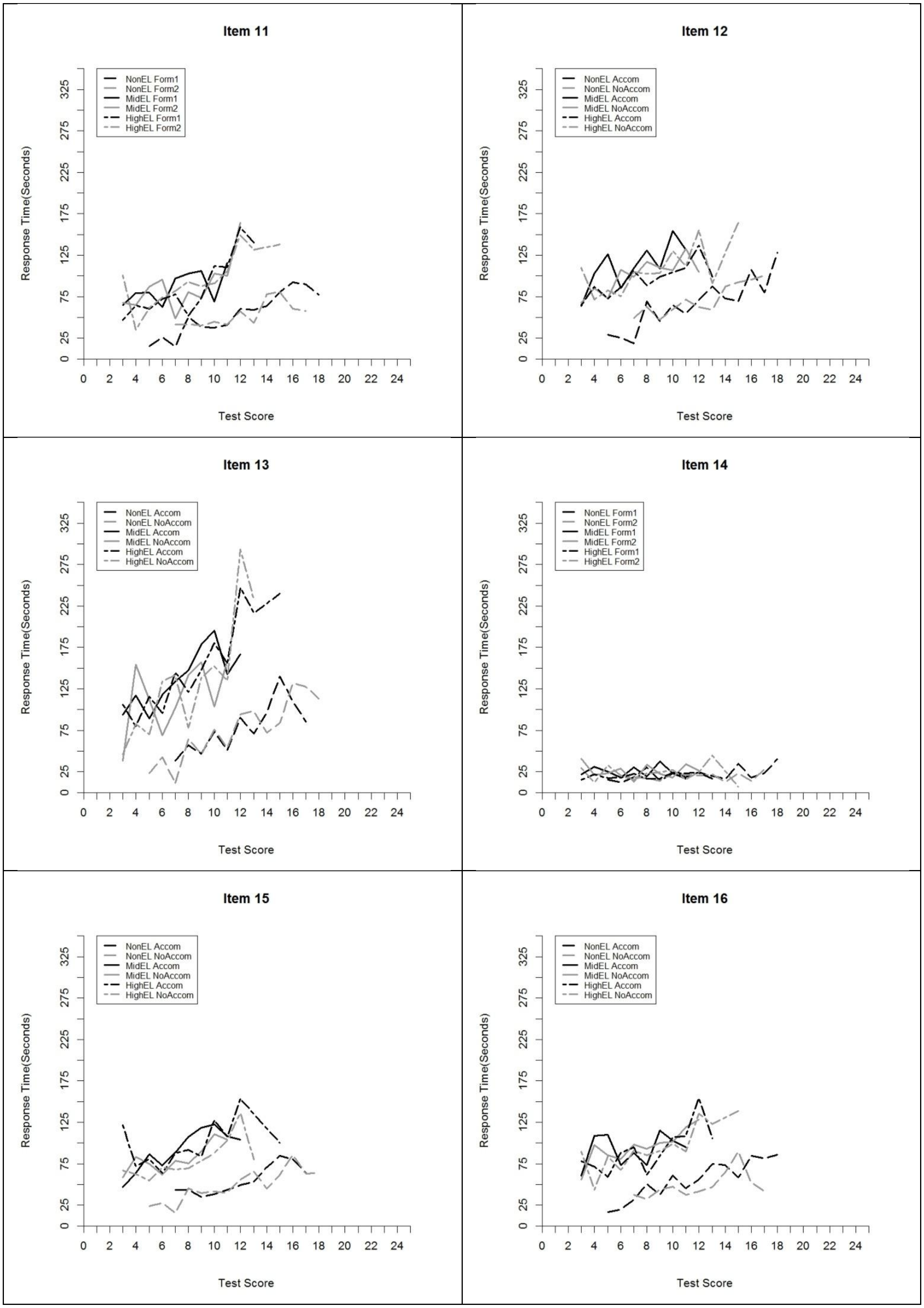
APPENDIX F

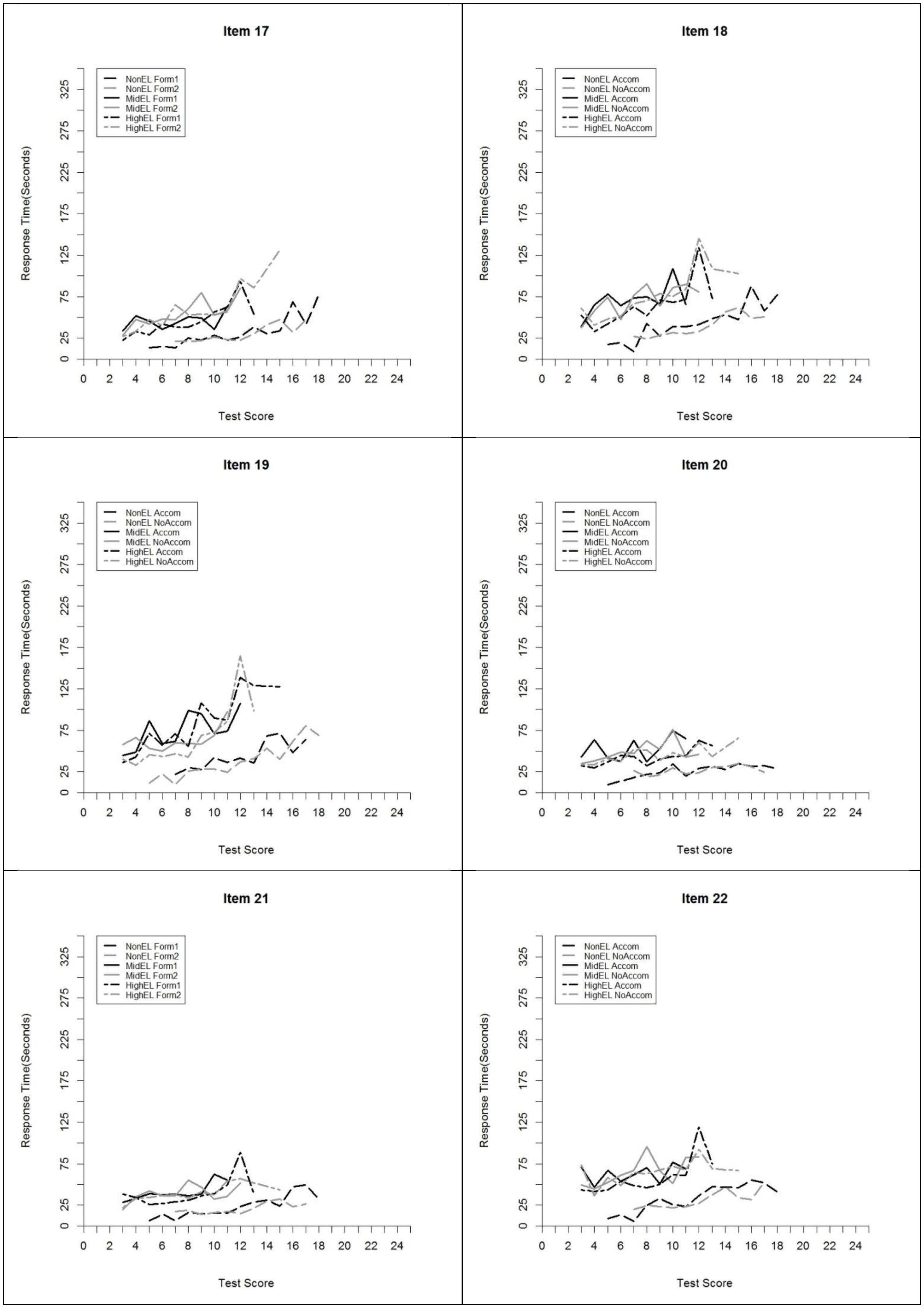
MATHEMATICS ITEM-LEVEL RESPONSE TIME (SECONDS) GRAPHS

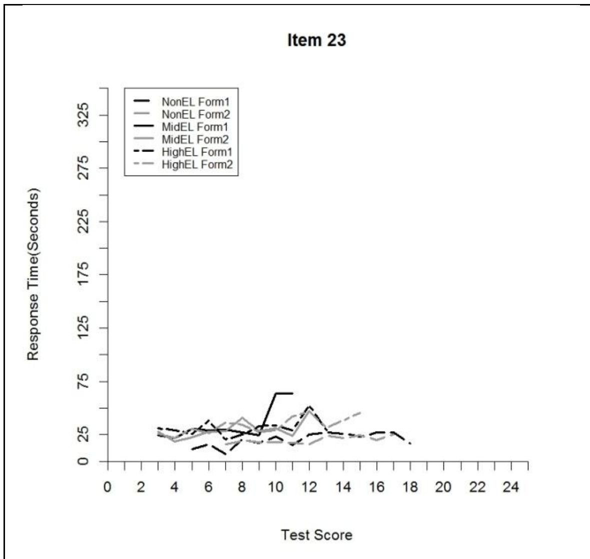


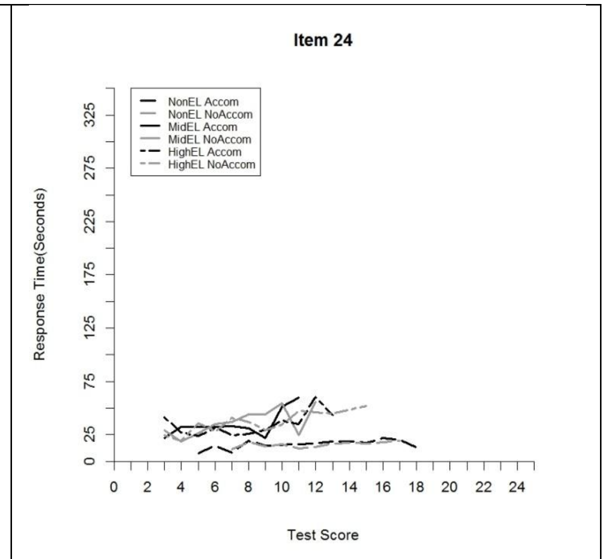


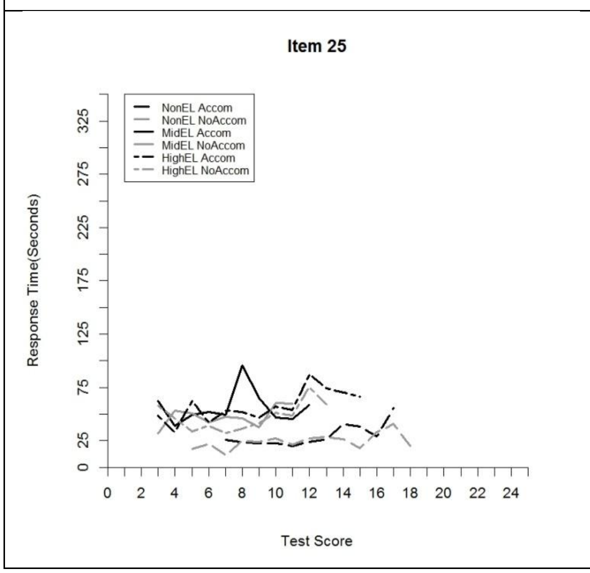
Item 5
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
Test Score
Item 6
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
Test Score
Item 7
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
Test Score
Item 8
NonEL Form1
NonEL Form2
MidEL Form1
MidEL Form2
HighEL Form1
HighEL Form2
Response Time(Seconds)
Test Score
Item 9
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
Test Score
Item 10
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
Test Score


Item 11
NonEL Form1
NonEL Form2
MidEL Form1
MidEL Form2
HighEL Form1
HighEL Form2
Response Time(Seconds)
0 25 75 125 175 225 275 325
0 2 4 6 8 10 12 14 16 18 20 22 24
Test Score
Item 12
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
0 25 75 125 175 225 275 325
0 2 4 6 8 10 12 14 16 18 20 22 24
Test Score
Item 13
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
0 25 75 125 175 225 275 325
0 2 4 6 8 10 12 14 16 18 20 22 24
Test Score
Item 14
NonEL Form1
NonEL Form2
MidEL Form1
MidEL Form2
HighEL Form1
HighEL Form2
Response Time(Seconds)
0 25 75 125 175 225 275 325
0 2 4 6 8 10 12 14 16 18 20 22 24
Test Score
Item 15
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
0 25 75 125 175 225 275 325
0 2 4 6 8 10 12 14 16 18 20 22 24
Test Score
Item 16
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
0 25 75 125 175 225 275 325
0 2 4 6 8 10 12 14 16 18 20 22 24
Test Score


Item 17
NonEL Form1
NonEL Form2
MidEL Form1
MidEL Form2
HighEL Form1
HighEL Form2
Response Time(Seconds)
Test Score
Item 18
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
Test Score
Item 19
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
Test Score
Item 20
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
Test Score
Item 21
NonEL Form1
NonEL Form2
MidEL Form1
MidEL Form2
HighEL Form1
HighEL Form2
Response Time(Seconds)
Test Score
Item 22
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
Test Score


Item 23
NonEL Form1
NonEL Form2
MidEL Form1
MidEL Form2
HighEL Form1
HighEL Form2
Response Time(Seconds)
0 25 75 125 175 225 275 325
0 2 4 6 8 10 12 14 16 18 20 22 24
Test Score


Item 24
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
0 25 75 125 175 225 275 325
0 2 4 6 8 10 12 14 16 18 20 22 24
Test Score


Item 25
NonEL Accom
NonEL NoAccom
MidEL Accom
MidEL NoAccom
HighEL Accom
HighEL NoAccom
Response Time(Seconds)
0 25 75 125 175 225 275 325
0 2 4 6 8 10 12 14 16 18 20 22 24
Test Score

REFERENCES

Abedi, J. (2001). *Assessment and accommodations for English language learners: Issues and recommendations* (CRESST Policy Brief 4). Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J. (2004). The no child left behind act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14.

Abedi, J. (2009). Computer testing as a form of accommodation for English learners. *Educational Assessment, 14*, 195-211.

Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2005). *Language accommodations for English language learners in large-scale assessments: Bilingual dictionaries and linguistic modification* (CSE Technical Report 666). Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, and University of California Los Angeles.

Abedi, J., Courtney, M., & Leon, S. (2003a). *Effectiveness and validity of accommodations for English language learners in large-scale assessments* (CSE Report 608). Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, and University of California Los Angeles.

Abedi, J., Courtney, M., & Leon, S. (2003b). *Research-supported accommodation for English language learners in NAEP* (CSE Rep. No.586) Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, and University of California Los Angeles.

Abedi, J. & Dietal, R. (2004). *Challenges in the No Child Left Behind Act for English language learners* (CRESST Policy Brief 7). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J. & Herman, J. (2010). Assessing English language learners' opportunity to learn mathematics: Issues and limitations. *Teachers College Record, 112*, 723-746.

Abedi, J., Hofstetter, C., Baker, E., & Lord, C. (2001). *NAEP math performance and test accommodations: Interactions with student language background* (CSE Tech. Rep. No. 536). Los Angeles. CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, and University of California Los Angeles.

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1-28.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3). 219-234.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, *19*(3), 16–26.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bayley, R., Abedi, J., & Ewers, N. (2010, May). *Providing computerized language-based accommodations in the assessment of English language learners*. Retrieved from http://slai.ucdavis.edu/sites/slai.ucdavis.edu/files/attachments/08_Bayley_RealCAGold.pdf

Beech, M. (2010). *Guide to FCAT and FCAT 2.0 accommodations for students with disabilities*. Tallahassee, FL: Bureau of Exceptional Education and Student Services, Florida Department of Education.

Bergstrom, B. A., Gershon, R., & Lunz, M. (1994, April). *Computer adaptive testing: Exploring examinee response time using hierarchical linear modeling*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Butler, F. A. & Stevens, R. (1997). *Accommodation strategies for English language learners on large-scale assessments: Student characteristics and other considerations* (CSE Technical Report 448). Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, and University of California Los Angeles.

Calhoon, M. B., Fuchs, L. S., & Hamlett, C. L. (2000). Effects of computer-based test accommodations on mathematics performance assessments for secondary students with learning disabilities. *Learning Disability Quarterly, 23*(4), 271-282.

California Department of Education. (2009). *California English language development test (CELDT: Understanding and using 2009-10 individual results*. Retrieved from http://www.cde.ca.gov/ta/tg/el/documents/celdt09astpkt1.pdf

Cawthon, S. W. (2010). Assessment accommodations for English language learners: The case of former-LEPs. *Practical Assessment, Research & Evaluation, 15*(13).

Chang, S., Plake, B. S., & Ferdous, A. A. (2005, April). *Response times for correct and incorrect item responses on computerized adaptive tests*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Cho, H., Lee, J., & Kingston, N. (2012). Examining the effectiveness of test accommodation using DIF and mixture IRT model. *Applied Measurement in Education, 25*(4), 281-304.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Collins, L. M. & Lanza, S. T. (2010). *Latent class analysis and latent transition analysis: With applications in the social, behavioral, and health sciences*. Hoboken, NJ: John Wiley & Sons, Inc.

Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying the principles of universal test design to delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *The Journal of Technology, Learning, and Assessment, 3*(7).

Forte, E. & Faulkner-Bond, M. (2010). *The administrator's guide to federal programs for English learners.* Washington, DC: Thompson Publishing Group, Inc.

Fry, R. (2008). *The role of schools in the English language learner achievement gap*. Washington, DC: Pew Hispanic Center. Retrieved from http://pewhispanic.org/files/reports/89.pdf

Galindo, C. (2009). *English language learners' math and reading achievement trajectories in the elementary grades: Full technical report*. Baltimore, MD: University of Maryland, Baltimore County.

Giraud, G. & Smith, R. (2005). The effect of item response time patterns on ability estimates in high stakes computer adaptive testing.

Gorin, J. S. (2006). Test design with cognition in mind. *Educational Measurement: Issues and Practice*, *25*(4), 21-35.

Gorman, S. (2010). *An introduction to NAEP* (NCES 2010-468). Washington, DC: National Center for Educational Statistics.

Gravetter, F. J. &Wallnau, L. B. (2002). *Essentials of statistics for the behavioral sciences* (4th Ed.). Pacific Grove, CA: Wadsworth Group.

Hakuta, K., Butler, Y., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* (Policy Report 2000-1). Berkeley, CA: University of California Linguistic Minority Research Group.

Hofstetter, C. H. (2003). Contextual and mathematics accommodation test effects of English-language learners. *Applied Measurement in Education, 16*(2), 159-188.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.

Hornke, L. F. (2000). Item response times in computerized adaptive testing. *Psicologica, 21*, 175-189.

Hornke, L. F. (2005). Response time in computer-aided testing: A "verbal memory" test for routes and maps. *Psychology Science, 47*(2), 280-293.

Jiao, H., Lissitz, R. W., Macready, G., Wang, S., & Liang, S. (2011). Exploring levels of performance using the mixture Rasch model for standard setting. *Psychological Test and Assessment Modeling*, *53*(4), 499-522.

Kellow, J. T. & Jones, B. D. (2008). The effects of stereotypes on the achievement gap: Reexamining the academic performance of African American high school students. *Journal of Black Psychology, 34*(1), 94-120.

Kieffer, M.J., Lesaux, N.K., Rivera, M., & Francis, D.J. (2009). Accommodations for English language learners taking large-scale assessments: a meta-analysis of effectiveness and validity. *Review of Educational Research, 79*(3), 1168-1201.

Kim, J. & Herman, J.L. (2009). A three-state study of English learner progress. *Educational Assessment, 14*, 212-231.

Koenig, J. A. & Bachman, L. F. (Eds.). (2004*). Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessments*. Washington, DC: National Academies Press.

Kopriva, R.J. (2008). *Improving testing for English language learners*. New York, NY: Taylor and Francis.

LaCelle-Peterson, M. W. & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review, 64*(1), 55-75.

Llabre, M. M. & Froman, T. W. (1987). Allocation of time to test items: A study of ethnic differences. *The Journal of Experimental Education, 55*(3), 137-140.

National Center for Educational Statistics. (2012). *English language learners in public schools*. Retrieved from http://nces.ed.gov/programs/coe/pdf/coe_ell.pdf

National Council of Teachers of English. (2008). *English language learners: An NCTE policy research brief*. Retrieved from http://www.ncte.org/library/NCTEFiles/Resources/PolicyResearch/ELLResearchBrief.pdf

NCEO. (2011, March). *Don't forget accommodations! Five questions to ask when moving to technology-based assessments* (NCEO Brief #1). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 1001, 115 Stat. 1440. (2002).

O'Neill, K. & Powers, D. E. (1993, April). *The performance of examinee subgroups on a computer-administered test of basic academic skills*. Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.

Parshall, C.G., Mittelholtz, D., & Miller, T. R. (1994, April). *Response time: An investigation into determinants of item-level timing*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Parshall, C. G., Davey, T., & Pashley, P. J. (2000). Innovative item types for computerized testing. In W. J. van der Linden & G. A. W. Glas (Eds.), *Computerized Adaptive Testing: Theory and Practice* (pp. 129-148). Dordrecht, Netherlands: Kluwer Academic Publishers.

Pennock-Roman, M. & Rivera, C. (2011). Mean effects of test accommodations for ELLs and non-ELLs: A meta-analysis of experimental studies. *Educational Measurement: Issues and Practice, 30*(3), 10-28.

Pitoniak, M. J., Young, J. W., Martiniello, N., King, T. C., Butuex, A., & Ginsburgh, M. (2009). *Guidelines for the assessment of English language learners*. Princeton, NJ: Educational Testing Service.

Piza, E. L. (2012). *Using poisson and negative binomial regression models to measure the influence of risk on crime incident counts.* Newark, NJ: Rutgers Center on Public Security.

Rea, L. M., & Parker, R. A. (1992). *Designing and conducting survey research*. San Francisco: Jossey-Boss.

Ready, D. & Tindal, G. (2006). *An investigation of language-minority children: Demographic characteristics, initial performance, and growth in achievement* (CSE Technical Report 686). Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, and University of California Los Angeles.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*(3), 271-282.

Rumberger, R.W. & Gandara, P. (2004). Seeking equity in education of California's English learners. *Teachers College Record, 106*(10), 2032-2056.

Russell, M., Kavanaugh, M., Masters, J., Higgins, J., & Hoffmann, T. (2009). Computer-based signing accommodations: Comparing a recorded human with an avatar. *Journal of Applied Testing Technology*, *10*(3).

Scarpati, S. E., Wells, C. S., Lewis, C., & Jirka, S. (2011). Accommodations and item-level analysis using mixture differential item functioning models. *The Journal of Special Education, 45*(1), 54-62.

Schaeffer, G. A., Reese, C. M., Steffen, M., McKinley, R. L., & Mills, C. N. (1993). *Field test of a computer-based GRE general test* (ETS Research Report 93-07). Princeton, NJ: Educational Testing Service.

Schnipke, D. L. & Scrams, D. J. (2002). Exploring issues of examinee behavior: Insights gained from response-time analysis. In C. N. Mills, M. T. Potenza, J. J. Fremer, & W. C. Ward (Eds.), *Computer-Based Testing: Building the Foundation for Future Assessments* (pp. 237-266) Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Schnipke, D. L. & Pashley, P. J. (1997, March). *Assessing subgroup differences in item response times.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Schnipke, D. L. (1995, April). *Assessing speededness in computer-based tests using item response times*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Scrams, D. J., & Schnipke, D. L. (1999). *Making use of response times in standardized tests: Are accuracy and speed measuring the same thing?* Newtown, PA: Law School Admission Council.

Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature* (Center for Educational Assessment Research Report no. 485). Amherst, MA: Center for Educational Assessment, University of Massachusetts Amherst.

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

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35,* 4-28.

Sullivan, P., Yeager, M., Chudowsky, N., Kober, N., O'Brien, E., & Gayler, K. (2005). *States try harder, but gaps persist: High school exit exams.* Washington, DC: Center on Education Policy.

Thissen, D. (1983). Timed testing: An approach using item response theory. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 179-203). New York: Academic Press.

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

Thompson, S., Thurlow, M., & Moore, M. (2003). *Using computer-based tests with students with disabilities* (Policy Directions No. 15). Minneapolis, MN: University of Minnesota, National Center for Educational Outcomes.

U.S. Department of Education. (2005). *How no child left behind benefits Hispanic Americans.* Retrieved from http://www2.ed.gov/nclb/accountability/achieve/nclb-hisp.pdf

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics, 31*(2), 181-204.

Verbic, S. & Tomic, B. (2009). *Test item response time and response likelihood*. Retrieved from http://arxiv.org/ftp/arxiv/papers/0901/0901.4356.pdf

Wolf, M.K., Kao, J.C., Griffin, N., Herman, J.L., Bachman, P.L., Chang, S.M., & Farnsworth, T. (2008). *Issues in assessing English language learners: English language proficiency measures and accommodation uses practice review* (CRESST Report 732). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Graduate School of Education and Information Studies, University of California Los Angeles.

Wright, B. D. & Masters, G. N. (1982*). Rating scale analysis: Rasch measurement*. Chicago, IL: MESA Press.

Yang, X., Shaftel, J., Glassnapp, D., & Poggio, J. (2005). Qualitative or quantitative differences? Latent class analysis of mathematical ability for special education students. *The Journal of Special Education, 38*(4), 194-207.

Zenisky, A. L. & Baldwin, P. (2006). *Using response time data in test development and validation: Research with beginning computer users* (Center for Educational Assessment Report No. 593). Amherst, MA: Center for Educational Assessment, University of Massachusetts Amherst.