Technical characteristics of e-based vs. paper-pencil CBM tasks for students who are deaf and hard of hearing

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Elizabeth Anne Lam

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Kristen McMaster, Co-Advisor
Dr. Susan Rose, Co-Advisor

December, 2015

Copyright Page

Acknowledgements

Success in my doctoral studies and this dissertation would not have been possible without the help and support from others. First, I would like to thank my husband, Dave, who supported me through inception, research design, data collection and analysis. His encouragement, belief in my abilities and support in managing the household when I needed to commit time to my studies was amazing. Second, I would like to thank my 18-month-old son, Evan. Evan was there to brighten my day, help me keep perspective, and remind me to delight in the amazing world around me. Third, I would like to thank my parents (Chuck and Linda), sister (Kathryn) and parents-in-law (Peter and Emerie), who cared for our family throughout the dissertation process. The dissertation would not have been possible without your support.

I would like to thank my dissertation advisors, Dr. McMaster and Dr. Rose. Your encouragement, high attention to detail, and expertise was critical the success of this dissertation. I would also like to thank Dr. Hooper at Penn State, who assisted me with the technological aspect of the research study. Thank you also to my dissertation committee, Dr. Christ and Dr. Nelson who provided additional insight in the area of measurement and deafness.

Next, I would like to thank the National Leadership Consortium in Sensory Disabilities (NLCSD) who made my doctoral studies possible through opportunities to engage in scholarly discourse, guidance and financial support. Thank you to Dr. Brooke Smith and Dr. Heather Hayes who encouraged me to grow professionally throughout this journey, university consortium members and to my fellow NLCSD fellows.

Last, I would like to thank the students, families and teachers who participated in this study. Without your commitment to generating research for the field of deafness, this dissertation would not be possible. I would also like to thank the researchers who will use findings from this dissertation to further research within the area of Curriculum-Based Measures (CBM) and deafness.

Abstract

To increase literacy outcomes for students who are deaf and hard of hearing, professionals need to provide high quality instruction that is informed by accurate and valid assessment data. This study compared the reliability and validity of student scores from paper-pencil and e-based assessments, "Maze" and "Slash." Forty (N=40) students, who were deaf or hard of hearing and read between the second and fifth grade reading level, participated. Twenty-one teachers of students who are deaf and hard of hearing also participated. For Maze, alternate form reliability coefficients obtained from correct scores and correct scores adjusted for guessing ranged from .61 to .84 (ps < .01); criterion-related validity ranged from .33 to .67 (majority of ps < .01). These findings are generally consistent with findings from previous research. For Slash, alternate form reliability coefficients obtained from correct scores ranged from .50 to .75 (ps < .01); criterion-related validity ranged from .25 to .72. The extent to which testing modifications delivered in an electronic-based (e-based) format influenced student scores was also examined. Differences between paper-pencil and e-based conditions were generally non-significant for Maze; significant differences between conditions for Slash favored the paper-pencil condition. Overall, findings suggest that Maze holds promise for use with students who are deaf and hard of hearing in both conditions, with inconclusive results for Slash. Future research is needed to explore the impact of providing testing modifications through e-based progress monitoring tools.

Table of Contents

List of Tables

CHAPTER 1

INTRODUCTION

Students who are deaf or hard of hearing are a heterogeneous population, with some students meeting or exceeding English literacy proficiency benchmarks and others demonstrating limited proficiency (Gilbertson & Ferre, 2008). Researchers have long recognized a subset of students who are deaf and hard of hearing who demonstrate limited reading proficiency (Allen, 1986; Allinder & Eccarius, 1999; Karchmer & Mitchell, 2003; Luckner, 2013; Rose, 2007; Trybus & Karchmer, 1977). Limited English literacy proficiency can substantially reduce the opportunities available to students who are deaf or hard of hearing (Appelman, Callahan, Mayer, Luetke & Stryker, 2012; Boutin, 2008; Hartmann, 2010; Luckner, Sebald, Cooney, Young & Muir, 2005; Luft, 2012; Luft & Huff, 2011).

Providing high-quality instruction designed to meet students' individual needs in the school years will likely decrease the trend that the "average student with a hearing loss graduates from high school with reading comprehension skills at approximately the fourth-grade level" (Luckner, 2013, p. 8). There is a need for researchers to develop and practitioners to implement high-quality instructional approaches designed to address the persistent problem of underachievement for this subgroup of students (Easterbrooks & Stephenson, 2006; Gilbertson & Ferre, 2008; Luckner et al., 2005; Moores, 2008).

To provide high-quality instruction, teachers need to use valid and reliable assessment measures to identify whether the student is responding to instruction and to modify the instructional approach if the response is inadequate (Luckner & Bowen, 2006). Researchers have stressed the importance of using assessment data to inform instruction when teaching students who are deaf and hard of hearing (Luckner & Bowen, 2006, 2010; Rose, 2007).

Progress Monitoring

Preliminary research suggests that CBM may hold utility with students who are deaf and hard of hearing (Luckner & Bowen, 2006, 2010; Rose, 2007). Some teachers who work with students who are deaf and hard of hearing use progress monitoring data in addition to achievement tests to gain a comprehensive picture of the reading skills of their students. The National Center on Progress Monitoring (n.d.) defines progress monitoring as “a scientifically based practice that is used to assess students’ academic performance and to evaluate the effectiveness of instruction” (What is Progress Monitoring section, para. 1). Progress monitoring tools, such as CBM, provide practitioners with data to inform instructional decision-making (Fuchs, Fuchs & Compton, 2012; Wanzek & Vaughn, 2010).

CBM, as conceptualized by Deno (1985), is a process that can effectively provide “vital signs” of a student’s educational health. CBM was designed to measure students’ responsiveness to instruction and inform service delivery. Deno (1985) described five key characteristics of CBM: the measures should be (1) reliable and valid, (2) efficient (3) easy to use (4) easy to communicate the results, and (5) inexpensive. CBM are designed to be quick, efficient, and sensitive to growth over short periods of time. Dorn (2010) described formative assessment, such as CBM, as “one of the most powerful tools available to guide classroom decisions” (p. 325).

There is a critical need for assessment tools that can serve as indicators of students’ educational health, are sensitive to growth over short periods of time, and can inform instructional decision-making for students who are deaf or hard of hearing. When practitioners who work with hearing students use CBM data to make instructional decisions, improved student literacy outcomes have been observed (Fuchs et al., 2012; Fuchs & Vaughn, 2012; Wanzek & Vaughn, 2010). The most commonly used tool used with hearing students is Oral Reading Fluency (students read a passage aloud with the correct number of words read in one min is recorded).

Need for Accommodations

Researchers suggest that students who are deaf and hard of hearing may decode and process written text differently from their hearing peers (Marschark, 2006; Marschark et al., 2011; Schirmer and McGough, 2005) and may require accommodations when engaging in assessments (Cawthon, 2011; Cawthon, & Leppo, 2013; Marschark, 2006; Rose, 2007). As such, when researchers advocate for the use of CBM with students who are deaf and hard of hearing, they also emphasize that the CBM tools used with hearing students would likely require accommodations when administered to students who are deaf and hard of hearing.

For example, oral reading fluency, initially designed for hearing students, require a heavy reliance on sequential decoding and an oral response. For hearing students, evidence suggests that decoding is a critical component of the reading process (Fuchs, Fuchs, Hosp, & Jenkins, 2001; National Reading Panel, 2000). As such, CBM relies on quick and accurate reading (decoding skills). Students who are deaf or hard of hearing may have no access or limited access to phonological information, and thus may use different avenues for decoding texts.

Additionally, students may struggle to clearly articulate an oral response or may need to translate the text from their primary code and mode of communication (American Sign Language) into English, which could increase the cognitive burden. Researchers also suggest that linguistic comprehension may hold greater relevance in the reading process than decoding skills for students who are deaf or hard of hearing (Mayberry, Del Giudice, & Lieberman, 2010).

Types of Testing Accommodations

Researchers have identified types of testing accommodations that may be appropriate when administering CBM tasks to students who are deaf and hearing, The suggested accommodations generally fall into four categories: directions and practice, student response, scoring, and test environment. For directions and practice, accommodations to promote access may include: presenting the directions in American Sign Language or modified modalities (e.g. Cued Speech, Visual Phonics, sign supported English), rephrased, repeated, demonstrated or illustrated directions, and additional opportunities to practice (Luckner & Bowen, 2010; Luckner, 2013). For student response, accommodation to promote access may include allowing students to respond in their primary mode of communication. For scoring, researchers recommend videotaping the testing session or administering the assessment with two examiners.

For testing environment, students who are deaf and hard of hearing may benefit from testing environments that are "executively controlled" (Gioia, Isquith, & Guy, 2001) with structure and modification provided to support students' engagement with the tasks (Hintermair, 2013; Kronenberger, Pisoni, Henning, & Colson, 2013; Marschark & Knoors, 2012; Oberg & Lukomski, 2011; Rhine, 2004). Testing accommodations may include supports in the areas inhibit impulses, shift attention, working memory, initiating tasks and monitoring performance.

Practitioners who currently administer CBM to students who are deaf and hard of hearing describe how providing these accommodations, though needed, require higher levels of staff time and school resources. Recommendations of how to provide these accommodations in a standardized way to reduce unintended variability in examiner delivery are also needed.

Using Technology to Deliver Testing Accommodations

The use of technology to promote test access is well documented in the literature for students without hearing loss (Dolan, 2005; Ketterlin-Geller, 2005; Mislevy, 2014; Rose, 2000; Russell, Hoffmann & Higgins, 2009; Salend, 2009). In the area of deafness, technology is frequently used to provide access (e.g. videophone, closed captioning) but limited research has explored how technology can be used to deliver assessment tools for students who are deaf and hard of hearing.

Purpose of the study

Preliminary research suggests that CBM appears to hold promise for students who are deaf and hard of hearing, with an understanding that testing accommodations are needed. Recommended testing accommodations have been identified in the literature, but solutions on how to address concerns related to how barriers in administration may impact student scores requires further exploration.

Research is needed to explore how to administer CBM to students who are deaf and hard of hearing in a way that provides students with the needed accommodations in a standardized way. In this study, I explore one possible solution of using electronic-based (e-based) administration as a tool for test administration. This study compared the reliability and criterion-related validity of student scores from CBM (Maze and Slash) administered in paper-pencil and e-based formats. It also compared student performance between the paper-pencil and e-based conditions.

This study used Avenue: PM (2011) as the e-based format. Avenue: PM (2011) is a suite of CBM e-based progress monitoring tools that provides a set of standardized accommodations designed for use with students who are deaf and hard of hearing. Avenue: PM addresses the testing accommodations suggested in the literature in four areas: directions and practice, student response, scoring, and test environment.

This study adds to the literature in two main ways. First, the study builds upon existing research by examining whether scores from CBM tools are consistent across leveled reading passages (reliability) and if the scores can serve as general outcome measures and predict student performance on criterion-referenced tests (criterion-related validity) for students who are deaf and hard of hearing. Second, this study is the first study to date to systematically compare student performance on paper-pencil versus e-based formats.

Research Questions

1. Does the Maze produce scores with sufficient alternate form reliability ($r \geq .80$) and sufficient criterion-related validity ($r \geq .50$, except IMC, r's ≥ -.50) on two administrative conditions (paper-pencil and e-based) for students who are deaf or hard of hearing?
2. Does the Slash produce scores with sufficient alternate form reliability ($r \geq .80$) and sufficient criterion-related validity ($r \geq .50$, except Total Wrong, r's ≥ -.50)

on two administrative conditions (paper-pencil and e-based) for students who are deaf or hard of hearing?

3. Do students who are deaf or hard of hearing perform reliably differently on the Maze when it is administered in paper-pencil versus e-based conditions?
4. Do students who are deaf or hard of hearing perform reliably differently on the Slash when it is administered in paper-pencil versus e-based conditions?
5. What is the feasibility of administering the paper-pencil and e-based conditions in the educational setting?

CHAPTER 2

LITEREATURE REVIEW

Researchers have long recognized a subset of students who are deaf and hard of hearing who demonstrate limited reading proficiency has been recognized in the literature for over 40 years (Allen, 1986; Allinder & Eccarius, 1999; Karchmer & Mitchell, 2003; Luckner, 2013; Rose, 2007; Trybus & Karchmer, 1977). Limited English literacy proficiency can substantially reduce the opportunities available to students who are deaf or hard of hearing in completing high school, succeeding in post-secondary education and successfully navigating employment settings (Appelman, Callahan, Mayer, Luetke & Stryker, 2012; Boutin, 2008; Hartmann, 2010; Luckner, Sebald, Cooney, Young & Muir, 2005; Luft, 2012; Luft & Huff, 2011).

To provide high quality instruction to combat underachievement, teachers need assessment tools that yield valid and reliable information about student performance (Luckner & Bowen, 2006, 2010; Rose, 2007). Standardized achievement tests are commonly used; however, researchers have raised concerns regarding using these tools with students who are deaf and hard of hearing (Cawthon, 2011; Cawthon, & Leppo, 2013; Marschark, 2006). Additionally, a majority of these tests have limited utility for instructional planning with students who are deaf and hard of hearing (Rose, 2007). Progress monitoring assessments, specifically curriculum-based measures (CBM), may serve as useful tools to complement achievement tests. CBM are defined as formative assessment tools that are: (1) reliable and valid, (2) efficient (3) easy to use (4) easy to clearly communicate the results, (5) inexpensive, (6) sensitive to student growth and (7) can be administered frequently over short periods of time (Deno, 1985).

Two highly relevant reviews have explored reading fluency and progress monitoring with students who are deaf and hard of hearing. Rose (2007) presented a conceptual review that described the utility of progress monitoring for students who are deaf and hard of hearing. Rose (2007) also explained how many students who are deaf and hard of hearing perform below grade level expectations, described the landscape of commonly used assessment practices, and discussed how progress monitoring could complement existing practices. Research to date that explored the utility of CBM with students who were deaf and hard of hearing and a case example was also provided.

Luckner and Urbach (2012) conducted a systematic review of the literature from 1970 to 2009 and identified studies that explored reading fluency; many of these studies included the use of CBM. Six studies were identified: four provided an intervention involving repeated reading whereas two did not provide an intervention. None of the studies met the quality indicators set forth by What Works Clearing House or by the U.S. Department of Education (2012). Both relevant reviews suggested that CBM appears to be a viable tool to use with students who are deaf and hearing. Modifications to the testing environment have been suggested in the literature, but further research is needed to explore the effects of testing modifications on student performance.

Researchers are beginning to explore how technology could be used to modify the testing environment. The National Center on Educational Outcomes (NCEO) explored how "technology enhanced assessments offers the opportunity to provide standardized sign support delivered through the test platform" (Shyyan, Christensen, Rogers, & Kincaid, 2014 Executive Summary, p. v). Shyyan et al. (2014) presented students who are deaf or hard of hearing with videos of signers translating test directions, and students rated their preferences. Shyyan et al. (2012) study provided an example of how modifications appropriate for students who are deaf and hard of hearing could be provided in a standardized way when using technology.

Rose and Dolan (2000) describe that formative comprehension checks that can be delivered immediately in an e-based environment in ways that "function less like the traditional test, and more like scaffolds with feedback. In this manner, they are more strategically useful to reading and provide support for building meta-awareness and self-monitoring strategies…" (p. 50). In this vein, Emary (2012) described how the Test Learning Method of providing immediate feedback in an e-based format could be appropriate for use with students who are deaf and hard of hearing.

For hearing students, "game-based learning," coined by Prensky (2003), has been found to "not only improve their motivational levels of students for taking a test, but can also improve their overall cognition, which results in better test performance (Dennis, 2013, p. 225). This Learning via Games e-based module technique was also proposed as an approach that may benefit students who are deaf and hard of hearing (Emary, 2012).

To explore the effects of the utility of e-based platforms, I first searched the literature to identify research studies where CBM was delivered to students who are deaf and hard of hearing in an e-based format. No studies were identified. As such, this literature review only includes research studies with CBM tasks delivered in the paper-pencil format to students who are deaf and hard of hearing. Of the studies identified, some researchers employed some of the testing modifications discussed earlier (e.g. directions, practice, student response, scoring, testing environment). For this literature review, I systematically explore whether scores from paper-pencil CBM assessments demonstrate properties of reliability and criterion-related validity. Second, I document whether testing modifications were used when delivering these tools.

Literature Search Procedures

To identify relevant studies, I first broadly searched the Academic Search Premier, ERIC, Education Index Retrospective (1929-1983), and Psych Info databases. Keyword terms related to deafness (*deaf, deafness, hearing impairment, hearing impaired, hearing loss, partial hearing,* and *hard of hearing)* were paired with CBM terms (*CBM, curriculum-based assessment, progress monitoring, response to intervention, cloze procedure, DIBELS, oral reading fluency, and miscue analysis*). Second, I conducted ancestral searches of two highly relevant conceptual/review articles (Luckner, 2012; Rose, 2007). Third, I identified the most frequently occurring data sources in my initial searches (*Exceptional Children, American Annals of the Deaf, Journal of Deaf Studies and Deaf Education,* and *University of Minnesota dissertations*) and examined current abstracts over the last five years for relevance. Last, I examined technical reports and studies from relevant progress monitoring websites: *Research Institute on Progress Monitoring* and the *National Center on Student Progress Monitoring.*

To be included in this review, the study needed to meet three criteria. First, each study needed to assess the reading performance of students who were deaf or hard of hearing using at least one CBM (or CBM-like) reading task. Studies were selected if they followed the CBM guidelines established by Deno (1985). Researchers in the field of deafness have suggested that more time may be required to score CBM when translation into Manually-Coded English or American Sign Language is needed (Luckner, 2013). As such, studies were included in this systematic review even if the researchers allocated more time to training and score interpretation than is typically described in the CBM guidelines (termed CBM-like).

Second, the study had to explore whether student scores from CBM tasks were reliable and valid when used with this population. Intervention studies using CBM as the dependent variable in an intervention study with students who are deaf and hard of hearing were excluded as student scores may have been influenced by the presence of the manipulated independent variable (intervention) (e.g., Enns & Lafond, 2007; Schirmer, Therrien, Schaffer, & Schirmer, 2009).

Third, the review was inclusive of peer-reviewed articles, technical reports, dissertations, and master's theses but needed to be written in English. This broad acceptance of many data sources was chosen due to the limited research published within this field of study.

Definition and Terms

For this analysis, reliability (e.g., test-retest, alternate form, inter-rater) and criterion-related validity of student performance scores when using CBM tools with students who are deaf and hard of hearing were explored. According to the American Educational Research Association (AERA, 1999), reliability refers to "the consistency of such measurements when the testing procedure is repeated on a population of individuals or groups" (p. 25). Reliability coefficients reported for each study are displayed in Table 1.

Table 1

Reliability of Student Performance Scores when using CBM

Citation	Analysis	Scores	Findings (*p* values reported when available)
		Sign Reading Fluency	
Allinder & Eccarius (1999)	Inter-rater	Words read correctly	40% to 100% (*M*= 79%)
		Idea units retold	0% to 100% (*M* =79%)
	Internal Consistency	1 min & 3 min passages	.89 to .97
	Alternate Form	One min probes	.85
		Three min probes	.94
Easterbrooks and Huston (2008)	Inter-rater	Fluency Envelope	.98
		Visual Grammar	.75
	Internal Consistency		.86
		Slash	
Rose et al. (2008)	Inter-rater	SRFT Form A, SRFT Form B, TOSCRF	92%
	Alternate Form	SRFT Form A & SRFT Form B	.92
		Cloze	
LaSasso (1980)	Internal Consistency	5th grade passages (six forms)	.67 to .82
Kelly & Ewoldt (1984)	Inter-rater (sample of 100 responses)	Meaningful to passage	82%
		Meaningful in sentence	81%
		Related to English form	79%
		Sign form classification	82%
		Maze	
Reynolds (1985)	Not calculated	NA	NA
Chen (2002)	Inter-rater	Maze	100%
	Alternate Form	Correct	.86 & .77
		Corrected	.85 & .79
		Scan	.78 & .62
		Incorrect	.55 &. 30
	Test-retest	Correct	.83 (p <. 01)
		Corrected	.85 (p <. 01)
		Scan	.82 (p <. 01)
		Accuracy	.39 (p <. 01)
		Incorrect	.11 (*ns*)
Devenow (2003)	Alternate Form	Correct (Phase 2)	.60 to .80 (p<. 001)
		Corrected (Phase 2)	.64 to .82 (p<. 001)
		Scan (Phase 2)	.45 to .70 (p<. 001)
Barkmeier & Rose (2009)	Alternate Form	Form A	.42 to .75
		Form D	.80 to .86
		Form A & Form D	.41 to .90
		Form E	-.21to .85

Note: SRFT= Silent Reading Fluency Test; TOSCRF= Test of Silent Contextual Reading Fluency

According to AERA (1999), validity refers to "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of the test" (p. 9). Further, "analyses of the relationship of test scores to variables external to the test provide [an] important course of validity evidence (AERA, 1999, p. 13). In the context of this review, researchers explored the strength of the relationship from student scores between the CBM (or CBM-like) tasks and other indicators of reading (e.g., achievement tests, teacher ratings). Table 2 describes the validity findings for each study by describing the type and name of criterion measure, scores, subtests and/or grade level clusters, the relationship between the CBM and criterion measure, and significance level.

Table 2

Criterion-Related Validity of Student Performance Scores when using CBM

Citation	Type of Measure	Name of Measure	Scores, subtests, or grade level	Relationship between measures	Significance
			Signed Reading Fluency		
Allinder & Eccarius (1999)	Achievement Test	TERA-DHH	*M* no. of words read one min	.30	*ns*
			M no. of words read three min	.21	*ns*
			M no. of idea units retold	.36	*ns*
			M no. of words retold	.46	*p*<.05
			M no. of unique words retold	.47	*p*<.05
			% of content words retold	.46	*p*<.05
Easterbrooks and Huston (2008)	Achievement Test	WRMT-R	Word Comprehension	.38 to .46	*p*<.05
			Passage Comprehension	.55 to .64	*p*<.01
			Total Comprehension	.43 to .50	*p*<.05 or *p*<.01
			Slash		
Rose et al. (2008)	Fluency Test	TOSCRF	SRFT-Form A	.84	Not Reported
			SRFT-Form B	.90	Not Reported
	Achievement Test	MAP	Elementary	.75 & .74	Not Reported
			Middle School	.62 & .69	Not Reported
			High School	.59 & .58	Not Reported
	Informal	Teacher Ratings	Elementary	.74 to .85	Not Reported
			Middle School	.60 to .69	Not Reported
			High School	.54 to .60	Not Reported
			Cloze		
LaSasso (1980)	CBM	Reading for Concepts series	Passages (3rd, 5th, & 7th)	Ranking (easy to difficult): 5th, 7th, 3rd grade	NA
Kelly & Ewoldt, (1984)	Achievement Test	SAT-HI	Verbatim	69%	*p*=.003
			Meaningful in passage	69%	*p*=.003
			Acceptable English form	49%	*p*<.001
	Story Retell	Story Retell	Verbatim	49%	*ns*
			Meaningful in passage	67%	*p*=.05
			Acceptable English form	53%	*ns*

			MAZE		
Reynolds (1985)	Achievement Test	Cooperative Reading Comprehension Test	Total Score	.69	Not Reported
	Achievement Test	Gallaudet Reading Test	Total Score	.62	Not Reported
Chen (2002)	Achievement Test	TOWL-3 (Winter, Spring)	Correct	.76 & .88	*p*<.01
			Corrected	.77 & .89	*p*<.01
			Scan	.82 & .83	*p*<.01
			Accuracy	.52 & .62	*p*<.05 & *p*<.01
			Incorrect	.26 & -.17	*ns*
	Informal Measure	Teacher Ratings (Winter, Spring)	Correct	.79 & .76	*p*<.01
			Corrected	.82 & .74	*p*<.01
			Scan	.80 & .74	*p*<.01
			Accuracy	.48 & .50	*p*<.05 & *p*<.01
			Incorrect	.19 & .01	*ns*
Devenow (2003)	Achievement Test	SAT	P1-2M Correct	.72 & .74	*p* <.05
			P1-2M Corrected	.74 & .75	*p*<.05
			P1-2M Scan	.56 & .71	*p*<.05
			P1-4M Correct	.64 & .64	*p*<.05
			P1-4M Corrected	.64 & .66	*p*<.05
			P1-4M Scan	.46 & .56	*p*<.05
			P1-Untimed Correct	.62 & .57	*p*<.05
			P1-Untimed Corrected	.60 & .60	*p*<.05
			P1-Untimed Scan	.02 & .49	*ns* & *p*<.05
			P2 Correct	.87 & .88	*p*<.001
			P2 Corrected	.89 & .89	*p*<.001
			P2 Scan	.77 & .83	*p*<.001
Barkmeier & Rose (2009)	Achievement Test	MAP	Elementary	.80 to .91	Not Reported
			Middle School	.59 to .91	Not Reported
			High School	-.10 to .85	Not Reported
	Informal Measure	Teacher Ratings	Elementary	.86 & 81	Not Reported
			Middle School	.86 & .85	Not Reported
			High School	-.20 & -.28	Not Reported

Note: TERA-DHH = Test of Early Reading Ability – Deaf or Hard of Hearing; WRMT-R = Woodcock Reading Mastery Test – Revised; TOSCRF= Test of Silent Contextual Reading Fluency; SRFT= Silent Reading Fluency Test; Correct = Correct MAZE choices; Corrected = Corrected MAZE choices; Incorrect = Incorrect MAZE choices; P1-2M =Phase 1 – Two min, P1-4M =Phase 1- Four Min, P1-Untimed = Phase 1 Untimed; P2= Phase 2; MAP= Measures of Academic Progress; CBM= Curriculum-based measurement; SAT-HI = Stanford Achievement Test – Hearing Impaired; TOWL-3 = Test of Written Language –Third Edition; SAT =Stanford Achievement Test

For the purpose of this review, descriptive classifications are used to denote the strength of the correlations. Because what constitutes “sufficient” and “insufficient” correlations are context dependent, I considered previous narrative reviews exploring how student scores functioned when administered CBM tasks with hearing students (Marston, 1989; Wayman et al., 2007) and recommendations from the field of measurement (Cohen, 1988; Cohen, 1992; Thorndike & Thorndike-Christ, 2010). According to these recommendations, for reliability, a correlation above .80 is considered "sufficient" and correlations below this level considered "insufficient." For validity, correlations above .50 are considered "sufficient" and correlations below this level considered "insufficient." These descriptive classifications should be interpreted with the understanding that standardized ranges have not been established for the deaf and hard of hearing population. To promote clarity, when the terms "sufficient" and "insufficent" are used within the text the accompanying correlation ranges are also reported.

Analysis of Studies

The results of this systematic search yielded nine studies, with each study exploring one of four CBM (or CBM-like) tools (sign reading fluency, Slash, Cloze, and Maze).

Sign reading fluency. Two research teams (Allinder & Eccarius, 1999; Easterbrooks & Huston, 2008) examined student scores from sign reading fluency yielded sufficient reliability and criterion-related validity. Allinder and Eccarius (1999) assessed prelingually deaf students with moderate to profound hearing loss who used an English-based signing system and attended general education elementary schools. The participants (*n*=36, ages 6 to 13) were administered the Comprehensive Reading Assessment Battery (CRAB; Fuchs, Fuchs, & Hamlett, 1989) and the Test of Early Reading Ability – Deaf or Hard of Hearing (TERA-DHH; Toubanos, 1995). Allinder &

Eccarius (1999) used signed directions with response modes that included fingerspelling, and signing. The sessions were videotaped and scored by two examiners.

Study results indicate that traditional scoring techniques of counting morphological endings as incorrect (verses not requiring these endings) yielded statistically significant poor performance ($p < .001$) for both the one-min and three-min timed samples. Inter-rater agreement varied substantially (0 to 100%), with mean agreement of 78%. Internal consistency and alternate form reliability correlations were sufficient ($r = .85$ to .97). For criterion validity evidence, correlations between the CRAB (Fuchs et al., 1989) and the TERA-DHH (Toubanos, 1995) were insufficient ($r = .21$ to .47).

Allinder and Eccarius (1999) required one-to-one correspondence between the printed and signed word; in contrast, Easterbrooks and Huston (2008) measured the quality of the rendered sign interpretation with a one-to-one correspondence not required. Easterbrooks and Huston (2008) assessed students with severe to profound bilateral sensorineural hearing loss who attended a school for the deaf ($n=29$, age 9 to 16). Sign-based communication (e.g., Signed English, Total Communication, American Sign Language) was used receptively by all participants and expressively by most participants. As a standardized measure of scoring sign reading fluency was not available for this population, the authors created the Signed Reading Fluency Rubric for Deaf Children (SRFT; as cited by Easterbrooks & Huston, 2008).

Study participants were first administered the book, *No No* (Cowley) from the Wright Group, and responses scored using the Signed Reading Fluency Rubric for Deaf Children and then, zero to four months later, were administered two subtests (Word

Comprehension and Passage Comprehension) from the Woodcock Reading Mastery Test - Revised (WRMT-R; Woodcock, 1987). Easterbrooks and Huston (2008) used testing modifications of: providing signed directions, accepting signed responses, videotaping the session, and re-watching the video when needed. Reliability (*r* = .75 to .98) and criterion-related validity with the WRMT-R (Woodcock, 1987) (*r*= .38 to .64) spanned the insufficient and sufficient ranges.

In both of the above studies, inter-rater agreement of students' signed response varied. In Allinder and Eccarius's (1999) study inter-rater agreement ranged from 0% to 100% (*M*=78%) and in Easterbrooks and Huston's (2008) study correlation coefficients were .98 for the fluency envelope and .75 for visual grammar. The difficulty in consistently and accurately scoring a student's signed response is consistent with the broader literature that highlights the difficulty in scoring students' signed responses.

A major limitation is that sign reading fluency requires translation from one language to a second language. Needing to translate a text while simultaneously reading the text is inherently a different task than a student reading the text in one language and responding in the same language (no translation required). A second major limitation is that American Sign Language and English do not have direct one-on-one correspondences, which leads to ambiguity in translation and scoring.

Slash. Slash is a less commonly used CBM reading tool. For this task, students respond by drawing a vertical line between letters to distinguish among words in sentences. Rose, McAnally, Barkmeier, Virnig and Long (2008) explored Slash with students who are deaf or hard of hearing. Rose et al.'s (2008) participants (*n*=101, grades 3 to 12) attended a residential school for the deaf with six students attending the community public school for part of the day. All students qualified for special education services due to their hearing loss status (*n*=36 mild to moderate, *n*=61 severe to profound, *n*=4 range unknown). Twenty-three percent of the sample had additional disabilities.

Rose and McAnally (2006) created the Silent Reading Fluency Test (SRFT; as cited by Rose et al., 2008) also termed "Slash." This measure used the formatting structure of the Test of Silent Contextual Reading Fluency (TOSCRF; Hammill, Wiederholt, & Allen, 2006) with content derived from Reading Milestones (Quigley, McAnally, Rose, & King, 2001) and Reading Bridge (Quigley, McAnally, Rose, & Payne, 2003). The Slash sentences were linguistically controlled using the developmental sequence of syntactical structures as identified by Quigley, McAnally, Rose, and King, (2001).

The SRFT and the TOSCRF, presented as three-min timed tasks, were administered within a three-day time frame quarterly throughout the academic year. The Measures of Academic Progress (MAP; NWEA, 2003) was administered twice a year, within 10 days of the progress monitoring measures. In the fall, teachers rated the student's reading ability (1 = insufficient general reading ability to 5= average general reading ability).

Inter-rater agreement (92%) and alternate form reliability (r = .92) were sufficient. For criterion-related validity, CBM scores were highly correlated with the TOSCRF (Hammill et al., 2006) (r = .84 and .90). The correlations between the SRFT and the MAP (NWEA, 2003) were in the sufficient range for all grade clusters (elementary r= .74 and .75, middle school r = .62 and .69, high school r = .58 and .59). Correlations between CBM and teacher ratings were in the sufficient range for elementary and middle school (elementary *r*= .85 and .82, middle school *r* = .51 and .58) and in the insufficient range for high school (*r* = -.34 and -.47).

The study had the following limitations. First, it is unknown if the limited readability range and controlled linguistic structures of the SRFT passages (first to fifth grade) impacted the correlations at the middle and high school levels. Second, the teacher ratings may have been enhanced if the teacher's rating of the participants' reading abilities were collected quarterly throughout the academic year and if the teachers' level of confidence (knowledge of the student's reading skill) was also documented. It is possible that teachers at the secondary level may have had less confidence in their ratings as students switched classes for content-area instruction (e.g., science, social studies, math, English). Despite these limitations, this study provides preliminary evidence of the utility of a Slash task with students who are deaf or hard of hearing.

Cloze. Two studies (Kelly & Ewoldt, 1984; LaSasso, 1980) explored the cloze procedure. LaSasso (1980) assessed prelingually profoundly deaf students who attended residential schools for the deaf (*n*= 95, ages 14 to 18). For this study, four passages were selected from the Reading for Concepts reading series (McGraw-Hill, 1970). The passages were selected at third, fifth, and seventh grade levels. Participants were administered three passages: third grade (selected from one of three forms), fifth grade (one of six forms), and seventh grade (one of three forms). LaSasso (1980) did not report the use of testing modifications to promote test access.

The internal consistency ranged from the insufficient to sufficient ranges (*r* = .67 to .82). LaSasso (1980) compared the mean student performance on the third grade, fifth grade, and seventh grade passages. Participants obtained the highest scores on the fifth-grade passage, the next highest score on the seventh-grade passage, and the lowest score on the third-grade passage. These findings were inconsistent with the expectation that participants would demonstrate the highest performance on the easiest passages (third-grade passages), moderate performance on the middle passage (fifth grade passage), and the lowest performance on the hardest passage (seventh grade passage).

The study had the following limitations. First, as multiple forms were created for each passage, fewer students were administered each passage. Because each cell size was small, extreme scores may have impacted the mean performance and subsequent results. Second, using a criterion-measure and correlation coefficients may have provided additional clarity on student performance. This study suggests student scores from paper-pencil CBM assessments had sufficient reliability but questions regarding validity when using this cloze response format remain.

Kelly and Ewoldt (1984) explored the cloze procedure when both verbatim responses and responses that were not verbatim but maintained text coherence were considered correct. Students who were deaf or hard of hearing and attended the Kendall Demonstration Elementary School (n=96, age 7 to 15) were assessed. Each student was administered a cloze passage at the student's reading level. Using a Borderline Group Technique, performance standards were established with the student performance on each of the three cloze scores classified as "acceptable" or "unacceptable." Additional data from a story retell (n=57) and performance on the Reading Comprehension subtest of the Stanford Achievement Test – Hearing Impaired (SAT-HI; as cited by Kelly & Ewoldt, 1984) (n=74) were analyzed using this same technique.

Inter-rater reliability conducted on 100 sample responses ranged from 79% to 82%. Using the consistency of decisions approach, the criterion validity of cloze passages with the SAT-HI (as cited by Kelly & Ewoldt, 1984) was 49% to 69%. Similar findings were reported for Story Retell.

The study had the following limitations. First, the demographic characteristics of the participant sample were not provided. Second, participants were administered one cloze passage that was scored by one rater. It is unknown if student performance was measure dependent or if there was any systematic variation in raters. Third, in addition to the consistency of decisions approach, the study would have been enhanced if the authors reported correlations between the scores on the cloze passage, story retell, and SAT-HI. This study highlights how considering verbatim and non-verbatim responses may be relevant when scoring cloze passages.

Maze. Four studies (Barkmeier & Rose, 2009; Chen, 2002; Devenow, 2003; Reynolds, 1985) explored the Maze, which is a modified-cloze technique (Fuchs & Fuchs, 1992). Reynolds (1985) assessed students who were deaf or hard of hearing who attended Gallaudet University. The participants (n=100, aged 18 to 30) ranged in hearing status from a 27 to 120 decibel loss (M = 93 decibel loss). The majority of the students were prelingually deaf (n=85). All students were assessed using the Degrees of Reading Power Test (DRP; College Board, 1980). Each passage consists of seven sentences with one word deleted. For each deleted word, five response choices were provided with one word accurately completing the sentence. Participants were administered 11 passages from the DRP (College Board, 1980). The students' Gallaudet admission battery data and demographic information were reported. Reynolds (1985) reported the use of a motivational feature (earning course credit points).

Students' DRP raw scores were converted to readability levels. These levels correspond to the level of text difficulty (e.g., children's magazine, general adult) that a student could comprehend at varying levels of accuracy (e.g., 50%, 75%, 90%). For this study, each student's readability level (level of prose text difficulty) at 75% accuracy was calculated and plotted in a group frequency distribution. Results yielded a bi-modal distribution with one peak at 55 to 59 (children's magazine readability) and another peak at 75-79 (professional journal readability). Although there was notable overlap between the performance of prelingual and postlingual subsamples, a statistically significant mean difference ($p < .001$) was present favoring the postlingual subsample.

Reliability was not reported. The criterion related validity between the Maze and Gallaudet admission measures was sufficient ($r = .62$ and .69). Correlations between the DRP scores and the demographic variables were also analyzed. Hearing status was non-significant predictor for the entire sample ($r = -.18$), but subgroup analysis yielded significance for the prelingual group ($r = -.26$, $p < .05$) but not for the postlingual group ($r = .21$). Articulation was significant predictor ($r = .54$, $p < .01$), but years of sign language experience ($r = -.10$) and speech reading ($r = .19$) were non-significant predictors.

The study had the following limitations. First, as demographic information was not provided separately for the prelingual ($n = 85$) and postlingual groups ($n=15$) it is unknown if the groups were comparable on other key variables (e.g., language development, presence of additional disabilities). Second, it is unknown if there was a systematic difference in participant characteristics between students who had admission data (55% of sample) and did not have admission data, which may have influenced study findings. Overall, these results suggest that the Maze procedure may hold promise for use with college students who are deaf or hard of hearing.

Chen (2002) explored CBM reading and writing passages with elementary students who are deaf or hard of hearing (age 6 to 12, *M*=10.12). All students attended the same school and received instruction either in a self-contained classroom (*n*=37) or in the mainstream (*n*=14). For the purposes of this review, emphasis is placed on the technical adequacy of Maze scores.

In this study, one-min Maze passages were administered in the winter (3 passages) and spring (3 passages). For the Maze, every seventh word is deleted and students select the best response given three choices. Five scores were obtained: Correct (number of correct choices selected), Corrected (number of correct choices minus number of incorrect choices divided by two), Scan (number of words viewed), Accuracy (percent of correct out of total number of choices) and Incorrect (number of incorrect choices selected). Students were also administered CBM written language probes, a subtest from the Test of Written Language –Third Edition (TOWL-3; Hammill & Larsen, 1996) and teacher ratings were obtained. Chen (2002) reported that the instructions were provided orally and in American Sign Language and students responded by circling their answer.

There was 100% inter-rater agreement. For alternate form and test re-test, CBM scores spanned the insufficient and sufficient ranges for the scores of Correct, Corrected and Scan (r = .62 to .86) and fell in the insufficient range for the scores of Accuracy and Incorrect (r = .11 to .55). For criterion-related validity, correlations between the Maze and the criterion measures (subtest from the TOWL-3 and Teacher ratings) were in the sufficient range for Correct, Corrected and Scan (r = .74 to .89) and fell in the insufficient and sufficient ranges for the remaining scores of Accuracy and Incorrect (r = -.17 to .62).

The study had the following limitations. The criterion measure was derived from only one subtest (TOWL-3; Hammill & Larsen, 1996) and was completed by a subset of participants (grades three to six) rather than the full sample. Despite these limitations, this study provides preliminary evidence supporting the utility of using the Maze with students who are deaf or hard of hearing.

Similar to Chen (2002), Devenow (2003) used the Maze but used a two-phase approach. Thirty-four students (ages 11 to 17, M=13.37) participated in Phase One and thirty-one students (ages 9 to 15, M=12.25) participated in Phase Two. All participants attended a residential school for the deaf either in the geographic Midwest (Phase One) or Southwest (Phase Two). American Sign Language was the primary mode of communication for the majority of participants with spoken English or a combination of both modes (speech and sign) also reported.

For Phase One, each participant was administered six Maze passages with two probes administered at each of the three testing conditions (untimed two-min, four-min). The scores of Correct, Corrected and Scan were used. The CBM scores in each of the three conditions were correlated with the participants' performance on the Stanford Achievement Test (SAT; Harcourt Brace Educational Measurement, 1995). For Phase Two, each participant was administered four Maze passages under the two-min condition.

Alternate form reliability, only calculated in Phase Two, ranged from insufficient to sufficient ranges (*r* = .45 to .80). For criterion-related validity, the correlations varied by administration length: two-min (sufficient range, *r* = .56 to .89), four-min (insufficient and sufficient ranges, *r* = .46 to .64), and untimed (insufficient and sufficient ranges *r* = .02 to .67).

The study had the following limitations. First, administration of the MAZE was less accurate when administered by teachers as compared to the primary researcher. In Phase One (teacher administered), 28 of the 62 participants were dropped from the study due to test administration errors. It is unknown if the dropped students were systematically different from the retained students, which could have impacted the results. In contrast, in Phase Two (researcher administered), 3 of the 34 participants were dropped due to administration errors. Second, since the criterion measure was only composed of one subtest, it is unknown if criterion validity would have remained constant with a composite reading score. Last, the sample size in this study was relatively small, which impacts statistical power and generalizability.

Using the same dataset as Rose et al. (2008), Barkmeier and Rose (2009) explored the Maze. Maze passages, derived from the Basic Academic Skills Sample (BASS), (Espin, Deno, Maruyama, & Cohen, 1989) were administered three times during the school year. The scoring criteria was the number of words incorrect subtracted from the total number of words correct (similar to the corrected choices score). Performance was compared to the MAP (NWEA, 2003) and teacher ratings.

Alternate form reliability varied substantially; correlations fell in the insufficient and sufficient ranges (*r* = -.21 to .90). For the MAP (NWEA, 2003) and teacher ratings, the correlations fell in the sufficient range for the elementary cluster (*r* = .80 to .91), sufficient range for the middle school cluster (*r* = .59 to .91) and spanned insufficient and sufficient ranges for the high school cluster (*r* = -.28 to .85).

The study had the following limitations. First, even though the sample size for the study was large (*n* = 101), when the sample was subdivided into elementary, middle school and high school, some grade clusters were small in size (elementary *n* =27, middle *n* = 16, high *n* = 58). Second, the accuracy of classroom teacher administration was not reported and thus the consistency of administration is unknown. Third, it is unknown if students scores on the MAP (NWEA, 2003) demonstrated sufficient reliability and criterion-related validity. Despite these limitations, the larger sample size presented in this study enhances the literature base on how student scores from Maze scores yield sufficient reliability and criterion-related validity. Across these four studies, results suggest that the Maze may hold utility with this population.

Study Comparison Across the Four CBM (or CBM-like) Types

Researchers' report of the reliability and criterion-related validity obtained from scores when students who are deaf and hard of hearing engaged in Sign Reading, Slash, Cloze and Maze are summarized below.

Reliability. Researchers reported reliability coefficients among different types (test-retest, alternate form, inter-rater) as generally sufficient with some variation. For Sign Reading Fluency, inter-rater agreement spanned the insufficient and sufficient ranges whereas internal consistency and alternate form fell in the sufficient range. These findings suggest that raters may have more difficulty efficiently and accurately recording a student's response when the response is in a signed form. The results of these two studies are consistent with the broader conversation in the field of deaf and hard of hearing that modification beyond simply changing the delivery mode from spoken English to sign language is needed when using CBM with this population.

For Slash, inter-rater agreement was above 90% and alternate form reliability was sufficient (*r* = .92). The symbol based response format coupled with clear guidelines on how to score if the vertical line is not clearly drawn, may have enhanced score consistency.

For Cloze, inter-rater agreement ranged from 79% to 82% and alternate form correlations spanned the insufficient and sufficient ranges (*r* = .67 to .82). For the cloze procedure, students read the passage and wrote in the missing words when they came to a blank line. Reliability was not consistently sufficient when the target word was required and was also not consistently sufficient when the target word was not required. It is possible that variability in the rater's ability to decipher student's written response and lack of clarity as to what constitutes an appropriate variation of the target word may have impacted reliability.

For Maze, reliability correlation coefficients appeared to vary based on the five scoring methods used: Correct (number of correct choices selected), Corrected (number correct after adjusting for guessing: Correct-Incorrect or Correct-Incorrect/2), Scan (number of words viewed), Accuracy (percent of correct out of total number of choices) and Incorrect (number of incorrect choices selected). For both alternate form and test-retest reliability types the strongest reliability coefficients were present for the scores of Correct and Corrected with variable and/or weaker correlations for the scores of Scan, Accuracy, and Incorrect.

Across all CBM and across the reliability types, the strongest reliability coefficients were noted for tasks where students provided a symbol based response (Maze and Slash), and more specifically for Maze with students who were in elementary school with the Maze scores of Correct and Corrected.

Validity. Validity results suggest that the strength of the correlations between the CBM (or CBM-like) tool and the criterion varied by the type of measure and score used. Two studies examined the validity evidence between sign reading fluency and an achievement test. Correlations fell in the insufficient to sufficient ranges when a one-on-one correspondence was and was not required (required r = .21 to .47, not required r = .38 to .64). Correlations between oral reading fluency and criterion measures with hearing students have been reported to range from .60 to .90 (Marston, 1989; Wayman et al., 2007), which is notably greater than the correlations presented in this review. These findings suggest that sign reading fluency with students who are deaf or hard of hearing does not relate to other criterion measures in the same way.

Slash was measured in one of the nine studies (Rose et al., 2008). The Slash was sufficiently correlated (r =.84 and .90) to the TOSCRF (Hammill et al., 2006). Preliminary research suggests that Slash may be a promising approach for students who are deaf or hard of hearing.

The cloze procedure was assessed in two of the nine studies. LaSasso (1980) noted that student performance on the leveled reading probes was inconsistent with the pre-established reading difficulty of each passage. Kelly and Ewoldt (1984) calculated the consistency of decisions approach with results for the achievement test and story retell falling between 49% and 69%. These preliminary results do not appear to suggest that this approach is an appropriate strategy for use with this population.

The Maze, assessed in four of the nine studies, was correlated to achievement test and/or teacher ratings. For college students, the correlations between the CBM and admission tests were sufficient (*r* = .62 and .69).

For school-aged children, four types of scores were used across the studies (Correct, Corrected, Scan, Accuracy, and Incorrect) and are described below. The correlations between Correct and the criterion measure were sufficient (*r* = .62 to .88) when the passages were administered in a one-min (Chen, 2002), two-min, four-min or untimed (Devenow, 2003) time frame. These results suggest that Correct appears to hold utility in measuring the Maze performance of students who are deaf or hard of hearing.

For Corrected, the manner in which this score was calculated varied across studies. Chen (2002) and Devenow (2003) calculated this score by taking the total number correct and subtracting it from the value of the number of incorrect by two. Barkmeier and Rose (2009) subtracted the number incorrect from the number correct but did not divide by two. Chen (2002) and Devenow (2003) correlations between Corrected and criterion measures were sufficient (*r* = .60 to .89). Barkmeier and Rose (2009) noted correlations falling in the sufficient range for the elementary clusters (*r* =.80 to .91) and the middle school cluster (*r* =.59 to .91) and in the insufficient and sufficient ranges for the high school (*r*= -.20 to .85) clusters.

Since both the elementary and middle school correlations in the Barkmeier and Rose (2009) study were generally consistent with the findings of Chen (2002) and Devenow (2003), it does not appear as if the differences in how errors were adjusted for consistently created variation between the findings; however, further research is needed to confirm this claim. Since Chen (2002) and Devenow (2003) did not segment the sample by grade, it is unknown if the correlations varied as a function of grade. Future research is needed to determine if the strength of the correlation between the criterion measure and the CBM, as measured by the Corrected, varies as a function of the students' grade level. Overall, these findings suggest that the Corrected CBM score may hold utility in measuring the performance of students who are deaf or hard of hearing especially for elementary and middle school students. Similar to the pattern observed for reliability, Correct and Corrected scores yielded the highest and most consistent correlations, as compared to Scan, Accuracy, and Incorrect.

The results across the nine studies suggest that although reliability was generally sufficient for most CBM tasks, the validity evidence supporting the use of CBM with students who are deaf or hard of hearing varied. Correlations for sign reading fluency with students who are deaf and hard of hearing was below correlations typically reported for oral reading fluency with hearing students. Slash appeared to hold promise for use with this population whereas the validity evidence supporting the use of the cloze procedure was weak. Validity evidence supported the use of the Maze with students who are deaf or hard of hearing especially for elementary and middle school students, when the task was one to two min in length, and the Correct and Corrected scores were employed. Overall, the validity evidence appeared the weakest for sign reading fluency and the cloze procedure and stronger for Slash and Maze. The results of the systematic review of the literature suggests that CBM, especially Maze and Slash, may hold promise when used with students who are deaf and hard of hearing.

Purpose of the Study

In this literature review, I examined student performance scores from CBM tasks when administered to students who are deaf and hard of hearing. All CBM tasks were presented using a paper-pencil format, and some researchers reported test modifications (e.g. signed directions, videotaping, second examiner).

Previous research suggests that using technology may be an appropriate option to provide testing accommodations to promote test access for students who are deaf and hard of hearing. This study is the first study to date to explore the effects of testing modifications using an e-based delivery format to deliver CBM.

The purpose of this study is twofold. First, I systematically explored whether Maze and Slash produce scores with sufficient reliability and criterion-related validity when administered to students who are deaf and hard of hearing. Second, I examined the effects of e-based testing modifications to Maze and Slash for students who are deaf and hard of hearing.

Research Questions

1. Does the Maze produce scores with sufficient alternate form reliability ($r \geq .80$) and sufficient criterion-related validity ($r \geq .50$, except IMC, r's $\geq -.50$) on two administrative conditions (paper-pencil and e-based) for students who are deaf or hard of hearing?
2. Does the Slash produce scores with sufficient alternate form reliability ($r \geq .80$) and sufficient criterion-related validity ($r \geq .50$, except Total Wrong, r's $\geq -.80$) on two administrative conditions (paper-pencil and e-based) for students who are deaf or hard of hearing?

3. Do students who are deaf or hard of hearing perform reliably differently on the Maze when it is administered in paper-pencil versus e-based conditions?
4. Do students who are deaf or hard of hearing perform reliably differently on the Slash when it is administered in paper-pencil versus e-based conditions?
5. What is the feasibility of administering the paper-pencil and e-based conditions in the educational setting?

CHAPTER 3

METHOD

This research study explored whether student scores on the Maze and Slash yield sufficient reliability and criterion-related validity when used with students who are deaf and hard of hearing. The study also explored the impact of the presence of the independent variable (administration type: paper-pencil vs. e-based) on student performance.

Below, I describe the setting, participating students who were deaf and hard of hearing, and participating teachers. Second, I describe the measures and materials used in this study, which include CBM Maze and Slash tasks, two criterion measures, and fidelity and utility ratings. Study procedures and data analyses are also described.

Setting and Participants

Due to the low incidence of deafness, recruitment efforts included multiple school districts. Below, I describe school district recruitment, teacher recruitment and selection, student eligibility criteria, student screening, and participant selection.

School district recruitment. Prior to implementation of this study, the University of Minnesota's Internal Review Board approved all procedures. I then engaged in three phases of recruitment: independent school districts, all intermediate school districts, and residential and charter schools for the deaf.

When recruiting at independent school districts, the goal was to contact schools with the highest number of students who were deaf or hard of hearing who met the eligibility criteria. I used a list of school districts in the state, which contained student count data. The top 20 school districts with the highest number of students were screened and removed from the list if: a) the district was served by an intermediate school (phase two recruitment would reach out to these districts), b) the external deadline for submissions to the district's research review board had passed, c) or the district was more than one and a half hours from the university.

After exclusion, seven school districts remained. I called and/or emailed a member of the deaf/hard of hearing team and/or the deaf or hard of hearing supervisor from the school district, most often the lead administrator. Of the seven districts, two districts never responded to the request, one district declined immediately, one district declined after viewing the study brochure, and two districts declined after the district's review process. One independent district elected to participate in the study and was approved by the school district's review board. After the student recruitment process (described below), I contacted the principal and classroom teacher for each eligible student.

For the second phase of recruitment, I reached out to administrators of the deaf and hard of hearing programs at all three intermediate school districts in the state. One district's administrator declined. For the remaining two districts, the administrators were interested and the proposal was approved. If a student was identified as eligible to participate (described later), I then contacted the member district to which the student attended. For each student, I contacted the student's principal and classroom teacher.

For the third phase of recruitment, two schools that served only students who were deaf and hard of hearing were contacted. One school did not respond to the request and the other school declined after a face-to-face meeting.

Following all recruitment including principal review, teacher review, parent consent and student assent (explained below), participating students (*n* = 40) attended 31 different schools. All schools were within a 25-mile radius of the university. One school was a private school; all others were public schools. In all schools, the primary language of instruction was spoken English.

Teacher recruitment and selection. All three districts (two intermediate districts and one independent district) employed teachers of deaf and hard of hearing (referred hereafter as "teachers"), who served students in the 31 schools. For each of the three districts, I was invited to the monthly teacher meeting where I proposed the research study to the teachers.

All teachers were given the opportunity to participate; having an eligible student on their caseload was not a requirement. The teachers were given three options for participation: (1) no involvement – did not provide consent, (2) low involvement, or (3) high involvement. The reason for the two options was that, during the development stage, one administrator indicated that given teachers' schedules and prior commitments, some teachers would likely be interested in participating but would be unable to commit to a high involvement option. For the low involvement option, teachers completed study forms and coordinated with the researcher to select appropriate testing times. For the high involvement option, teachers coordinated with the researcher throughout the study, engaged in training, obtained 90% accuracy in a practiced administration, administered all assessment tasks to the assigned students, and complete study forms.

Of the 23 invited teachers, two opted not to participate. The majority of the participating teachers were female (*n* = 20), Caucasian (*n* = 20) and non-Hispanic (*n* = 20). Two teachers reported having a hearing loss. The teachers' highest educational degree included: bachelor's (10%, *n* = 2), master's (33%, *n* = 12), and master's plus additional coursework (57%, *n* = 12). All teachers were licensed in the state of Minnesota as special education teacher of deaf and hard of hearing.

Approximately one-third of the teacher participants (38%) were licensed in an additional field (e.g. early childhood, speech and language pathologist, American Sign Language interpreter, secondary education-content, K-12 visual art specialist). The mean number of years in the current position was 11.48 (median = 7, range 1 to 40) while the mean total years teaching was 16.52 (median =12, range 1 to 40), and mean years as a teacher was 15.38 (median = 9, range 1 to 40). Of the 21 teachers who chose to participate, 11 selected the low involvement option and 10 selected the high involvement option.

Student eligibility criteria. After the introductory meeting, teachers provided me with a list of students on their caseload who met the student eligibility criteria. To be eligible to participate in the study the student needed to meet four criteria, (a) a documented hearing loss; (b) a reading level between the second to fifth grade according to teacher report, (c) placement in grades 2 through 12 and (d) no known motor or uncorrected vision impairment.

Student screening and selection. Teachers screened all students on their caseloads and identified 58 students who met the eligibility criteria. Three administrators (principals or directors of special services) declined participation due to a lack of interest or other district priorities (*n* = 4 students). Three classroom teachers (not the student's teacher of deaf and hard of hearing) declined participation of the three students in their respective classrooms due to not perceiving the study as the right fit given the student's educational needs or parent factors. With the removal of these seven students, 51 students were identified to engage in the parent consent process.

I introduced the study to the students, described the consent form, explained the importance of returning the form and the incentive for returning the form. I presented the consent and assent forms in spoken English, Total Communication or American Sign Language accommodating the students' communication modalities. (I am fluent in American Sign Language as evidenced by an Advanced Plus to Superior Plus Rating on the Sign Competency Proficiency Interview.)

The students took the consent forms home in their backpacks. If the form was not returned in one week, I provided the student with a duplicate paper copy to be sent home. As an incentive for returning the form, students received a gel pen.

I asked all participating teachers if any of the students or parents accessed print in a language other than English. If the language in the home was not English, I worked with the teacher to determine the best method to provide informed consent. I used interpreter services to relay the information verbally over the phone to the parents, conducted a home visit, and provided the consent form in the language accessible to the parents. Interpreter services were used to provide informed consent in the languages of Spanish and Hmong.

Of the 51 students, 42 parents provided consent, 3 parents denied consent, and 6 parents did not respond. When parent consent was received, the students were presented with an assent form. The assent form was presented in the student's preferred communication mode (e.g. Spoken English, Total Communication or American Sign Language). Forty students provided assent; 2 did not. Table 3 provides student demographic information including the number of students reported in each demographic category and the associated percentages. All participating student data were numerically coded and de-identified following test administration.

Table 3

Student Demographics

	All Students (n = 40)		Subset with MAP (n=27)	
Characteristic	Frequency	%	Frequency	%
Grade				
Elementary	33	83%	21	78%
Middle School	5	13%	4	15%
High School	2	5%	2	7%
Sex (M)	15	38%	11	41%
Ethnicity (Hispanic)	4	10%	3	11%
Race				
Asian	7	18%	7	26%
African American or Black	7	18%	6	22%
Caucasian	20	50%	10	37%
American Indian or Alaskan Native	0	0%	0	0%
Native Hawaiian or Other Pacific Islander	0	0%	0	0%
Other	6	15%	4	15%
Language				
ASL Only	1	3%	0	0%
Total Communication (Sign + Voice)	4	10%	1	4%
Spoken English Only	26	65%	18	67%
Not Listed	9	23%	8	30%
IEP Federal Setting				
1	35	88%	23	85%
2	4	10%	4	15%
3 & 4	1	3%	0	0%
Free/Reduced Lunch	16	40%	13	48%
Title 1	6	15%	5	19%
Hearing Loss				
Mild & Mild/Moderate	9	23%	5	19%
Moderate & Moderate/Severe	14	35%	10	37%
Severe & Profound	9	23%	4	15%
Unilateral (hearing loss in one ear only)	6	15%	6	22%
Sloping (hearing loss spans more than one range)	2	5%	2	7%
Amplification				
Cochlear Implant	8	20%	3	11%
Hearing Aid	22	55%	15	56%
Bone Conductor	2	5%	2	7%
Sound Field	4	10%	4	15%
None	4	10%	3	11%
Deaf Parent	2	5%	0	0%

Demographic information from this study was compared to the *Minnesota Department of Education: Students who are Deaf and Hard of Hearing* (2015) report. In Minnesota, there are a greater number of male students than female students receiving deaf and hard of hearing services (53% male, 47% female); the gender imbalance was reversed in this study (38% male). In Minnesota, the majority of students who receive deaf and hard of hearing services are White (65%). In this study there were fewer White students as compared to the state report (50%). In Minnesota, 74% of students spend at least 80% of their day in the regular class; in this study the percentage was slightly higher (88%). It appears as if the differences in this study to the state demographics may be attributed to the fact that students were not randomly selected from the Minnesota deaf and hard of hearing population and the sample size in this study is small (n=37).

Measures/Materials

The study included CBM (or CBM-like) tasks, criterion measures, and ratings of feasibility and utility. CBM-like refers to measures that generally follow the guidelines set forth by Deno (1985).

CBM and CBM-like probes. Eight unique third-grade one-min reading probes were used for this study. Four Maze passages were selected from the Children's Educational Services database and Edcheckup LLC (2005) and four Slash passages from the Test of Silent Contextual Reading Fluency (TOSCRF; Hammill, Wiederholt, & Allen, 2006) and Reading Milestones Placement and Monitoring test (RMPM, McAnally & Rose, 2012).

Maze passages. The content from Avenue: PM (2011) direction and practice item was used in the e-based condition and translated into the paper-pencil format (Appendix

A). The Maze passages were selected from the Children's Educational Services database and Edcheckup LLC (2005). The "Set 3" passages or third grade reading level passages, were reviewed. For inclusion, passages needed have a Lexile rating between 600 L and 730 L (25th to 75% quartile) and have a third grade Flesh-Kincaid reading level.

Of the 23 screened passages, 5 met the criteria. To fit within the character limitations of the e-based monitor display screen, these five passages were reduced in length to less than the 700 character limit. Each of the passages was re-checked to confirm that the reading level criteria were still met. One passage was ultimately dropped, as readability could not be successfully maintained while adhering to the character space limit guidelines. The four modified passages included in the study were: *The Accident* (paper-pencil and e-based), *Around the World* (e-based only), *Northern Pike* (paper-pencil only) and *The Race* (paper-pencil and e-based). The modified passages are provided in Appendices B, C, D, and E.

As distractors were randomly selected, each participant received the same passage content, and the same target word but with different distractors. As an example, in the passage "The Race" each student was given the same base sentence of "After **(___, ___, ___)** reading lesson was over, Mrs. Smith **(___, ___,___)** the class to the school playground." The target word was the same for each student ("the" and "led") but the target word varied in location (randomly assigned to position 1, 2 or 3) and the distractors were randomly selected from a database. To expand the example further, for student A the sentence read "After **(feels, cold, the)** reading lesson was over, Mrs. Smith **(for, led, snow)** the class to the school playground." For student B, the sentence read, "After **(since, hit, the)** reading lesson was over, Mrs. Smith **(few, led, egg)** the class to the school playground." This example provides an illustration of how each student received a standardized passage but not an identical passage.

Slash passages. The content from Avenue: PM (2011) direction and practice item was used in the e-based condition and translated into the paper-pencil format (Appendix F and G). To identify four Slash passages for this study, the passages from Avenue: PM (2011), Test of Silent Contextual Reading Fluency: Second Edition (TOSCRF: 2; Hammill, Wiederholt, & Allen, 2006), and Reading Milestones Placement and Monitoring (RMPM, McAnally & Rose, 2012) were reviewed. For inclusion of Slash passages, the Flesh-Kincaid Reading Level had to be at third grade, however a broader lexical range of 520L to 820L was used to account for variation in grammatical sequences.

Of the 261screened passages, 12 were identified as possible passages. These passages were then screened for the number of characters and the number of words in the passage. Upon review, four passages that most closely matched the readability criteria, were similar in character length, and had similar number of words in the passage were selected. Three passages were obtained from the TOSRF:2 and one passage from RMPM. To identify the passages, I labeled each passage by the first word of the passage. The Slash passages in the study are: "When" (paper-pencil and e-based), "Twelve" (paper-pencil and e-based), "The" (paper-pencil only) and "A" (e-based-only). See Appendix H, I, J, K for the content of each Slash passage. Unlike the Maze passages which had distractors that were randomized for each student, Slash passages were exactly the same for each student and across the paper-pencil and e-based condition.

Tasks. The two CBM or CBM-like tasks for this study were the Maze and Slash.

For the Maze, students were presented with a passage to read. The first sentence of the passage was left intact. For the remaining sentences in the passage, every seventh word was eliminated, replaced with a blank line, and three response choices provided. For paper-pencil, the target word and the two distractors were placed in parentheses with each word separated by a comma. All choices were printed in bold faced font. The target word was randomly assigned to the first, second, or third position. An example may read: "All three children started at the **(hers, since, same)** time and ran around the school." (Children's Educational Services database and Edcheckup LLC., 2005). In the e-based version, a student came to a blank line within the sentence and under the blank line was three words (one target word, two distractors) in parentheses in blue font. Similar to the paper-pencil version, the position of the target word (first, second or third) was randomized.

To complete the task in either condition, students read the passage and when they came to a parenthesis with three word choices or blank line (omitted word) they selected a word from multiple-choice options that completed the sentence. Students were able to change their answers at any time during the task.

If students finished before one min, the student would tell the administrator "done" and the administrator would record the time. To generate the students' prorated score, the student's obtained score was entered into the proration formula in Microsoft Excel. The proration formula of Prorated Score = 60 seconds/(Number of seconds used/recorded score) was used to calculate values of Correct Maze Choices and Incorrect Maze Choices. For example, if the student completed 18 correct items and 2 incorrect items in 45 seconds, to determine the number of Correct Maze Choices, the value of 18 was inputted into the “recorded score” place in the formula and 45 was inputted in the “number of seconds used” place in the formula. The formula calculated =60/(45/18) with a prorated score of 24 Correct Maze Choices. To calculate the Incorrect Maze choices, the same formula was used =60/(45/2) and the prorated value of 2.66 was obtained.

The Slash consisted of students marking boundaries between words when reading. More specifically, students were presented with a modified passage in which the story was presented in all upper case letters with no spaces or punctuation. When reading the passage, the students placed boundaries between the words with a slash mark. An example may read: “A G I R L R A N O U T O F T H E” and a student would draw a vertical line between the words: “A GIRL RAN OUT OF THE”. Similar to the Maze, if students made a mistake, they could change their answers.

Similar to the Maze, if the student finished before one min the score was prorated. Three prorated scores for Slash were generated: Correct Words Identified (paper-pencil only), Correct Boundaries (paper-pencil and e-based) and Total Wrong (paper-pencil and e-based). For example, if the student identified 23 words correctly, and drew 18 boundary lines and with one error in 45 seconds, the formula applied was: Correct Words Identified = 60/(45/23) =30.67, Correct Boundaries = 60/(45/18) = 24, and Total Wrong = 60/(45/1) = 1.33.

Scoring. Findings from a systematic review of the literature and the scoring features built into the Avenue: PM e-assessment were used as the basis for determining the scores. For the Maze, four scores were identified and used in this study: Correct Maze choices (CMC, total number of correctly identified selections within one min), Incorrect

Maze choices (IMC, total number of incorrect selections), Correct Maze Choices minus Incorrect Maze choices (CMC-IMC, total number of correctly identified selections minus incorrect selections within one min), and Correct Maze choices minus ½ Incorrect Maze choices (CMC- IMC/2, total number of correctly identified selections minus the total of incorrect selections divided by two).

For the Slash, only one technical report (Rose et al., 2008) explored the utility of this measure with students who are deaf and hard of hearing. The score used in Rose et al.'s (2008) technical report was the Correct Words Identified. For example, if the phrase " T H E D O G S L E E P S" was correctly completed and read " T H E ⁄D O G ⁄S L E E P S" the student would obtain a score of three, for three correctly identified words.

The scores generated by Avenue: PM (2011) were inconsistent with the scores used in the previous technical report (Rose et al., 2008), as the e-based management system scored Correct Boundaries rather than Correct Words Identified. In Avenue: PM (2011) three metrics were generated. First, the number of Correct Boundaries was the number of boundary lines selected by the student that correctly separated words within the probe. For example, in the phrase " T H E D O G S L E E P S" the total number of Correct Boundaries is two (THE⁄DOG⁄SLEEPS) as two lines were drawn and both lines correctly separated words within the phrase.

Second, the total number of Incorrect Boundaries was calculated as the summation of the number of boundary lines the student missed and the number of boundaries lines selected that did not correctly separate two words. For example, using the example above (T H E D O G⁄S L E E P⁄S) the total incorrect score was two: a missed boundary line between E and D (THE⁄DOG) and an incorrect boundary line placed between P and S.

Third, the Percent Correct was calculated by dividing the number of Correct Boundaries by the summation of Correct Boundaries and Incorrect Boundaries. For the e-based condition and paper-pencil conditions, the three scores described above (Correct Boundaries, Incorrect Boundaries, and Percent Correct) were used to score student responses. The traditional scoring method, Correct Words Identified, was quantified in the paper-pencil version.

Criterion measures. Two criterion measures were used for this study: MAP (NWEA, 2003) and Woodcock Johnson III Tests of Achievement Third Edition (WJ-III): Passage Comprehension subtest (Woodcock, McGrew, Mather, 2001).

When available, all participating students' Spring 2015 performance on the MAP (NWEA, 2003) reading test was obtained from the students' school districts. Of the study participants (full dataset), 68% (n = 27) participated in the MAP (NWEA, 2003). The MAP is an e-based adaptive assessment designed for second to twelfth grade students (NWEA, 2014). Students obtain a RIT score, which "represents the level of test item complexity at which he or she is capable of answering correctly about 50% of the time" (NWEA, 2014, p. 3). Test-re-test reliability with the 2002 NWEA norms in reading for third grade fall-to-spring was .87 and spring-to-spring was .89 (NWEA, 2004); however, there is no data related to the test-retest reliability with students who are deaf or hard of hearing.

Since the MAP (NWEA, 2003) is a commonly used standardized normed-referenced test that is designed to provide data as to how students are performing against the pre-established grade-level standard, it was selected for use in this study. As noted earlier, researchers in in the field of deafness note that standardized tests should be interpreted with caution when there are no data supporting the appropriateness of use with this population (Cawthon, 2011; Cawthon, & Leppo, 2013; Marschark, 2006; Rose, 2007). As such, an additional criterion measure was used to supplement the MAP (NWEA, 2003) for students who are deaf and hard of hearing.

Preliminary evidence from the Center on Literacy and Deafness (CLAD) indicates that the WJ-III Passage Comprehension subtest (Woodcock et al., 2001) may be appropriate for use with young students who are deaf or hard of hearing (Lederberg et al., 2014a, Ledererg et al., 2014b). As such, this measure was selected to gain additional criterion related validity evidence. The Woodcock-Johnson Test of Achievement: Fourth Edition (Schrank, Mather, & McGrew, 2014) was considered for use as it has recent norms; however, WJ-III Passage Comprehension subtest (Woodcock, et al., 2001) was ultimately selected as data from CLAD supports its appropriateness.

The WJ-III Passage Comprehension subtest (Woodcock et al., 2001), part of the larger WJ-III Test of Achievement, is an individually administered standardized test of achievement. This subtest measures reading and language comprehension through text (Woodcock et al., 2001). Students were presented with texts and prompted to identify the missing word in sentences and paragraphs. When administered to individuals without hearing loss, students respond orally and provided one-word answers (Schrank et al., 2001). The measure was hand scored. Raw scores were analyzed. Using the split-half procedure with the normative data, the median test reliability for the WJ-III Passage Comprehension subtest (Woodcock et al., 2001) is .80 with a median Standard Error of

Measurement (SEM) as 5.12 (Schrank et al., 2001).

When administered to students who are deaf and hard of hearing none of the stimulus materials (the sentences or passages) were translated into American Sign Language. However, I presented the directions in a language accessible to the student and translated any signed responses into English for scoring. If a student responded orally and the word was unintelligible the student was prompted to repeat the word in spoken English. If the student also used sign language, the student was asked to repeat the word in spoken English and/or fingerspell and/or sign the word. It is possible that these modifications to standardized administration and scoring could impact students' scores as neither of these criterion measures are designed specifically for use with students who are deaf or hard of hearing.

Feasibility and utility. Participating students and teachers rated feasibility and utility following administration of both conditions. For the student form, students rated which condition (paper-pencil or e-based) they "liked better" on three factors (directions, practice, and student response), which condition was "more fun," and which they would "recommend to a friend." Additionally, participating students rated for the e-based condition only, the level of helpfulness (not helpful, kind of helpful, or very helpful) for each of the following features: simulated directions, opportunity to re-watch the demonstration or re-try in the practice phase, feedback, text movement, changing answers, timer and motivational bar. If the students experienced difficulty understanding the feasibility and utility form, the student's teacher or I provided the needed clarification.

Following delivery of both conditions, the participating teachers completed the

Usage Rating Profile – Assessment (Chafouleas, Miller, Briesch, Neugebauer, & Riley-Tillman, 2012). Teachers completed this measure twice, once rating the paper-pencil condition and once rating the e-based condition. For each of the 28 items, teachers circled one of six ratings from "strongly disagree" to "strongly agree." According to Chafoules et al., (2012), the ratings from the items load onto six factors of: acceptability, understanding, home/school collaboration, feasibility, system climate, and system support.

Research Design

A correlational design was used to explore whether student scores on the Maze and Slash yield sufficient reliability and criterion-related validity when used with students who were deaf and hard of hearing. Additionally, the independent variable (administration type: paper-pencil vs. e-based) was manipulated to determine if students' scores from the Maze and Slash were reliably different across conditions. This section describes features of the research design and the next section explores how the independent variable (administration type: paper-pencil vs. e-based) was manipulated across conditions.

The repeated-measures research design was composed of three layers: (1) condition, (2) Maze and Slash, and (3) passages. These layers were used to ensure that differences between conditions were a result of true differences rather than another variable such as ordering effects. In this discussion, each layer was built upon until the full design was presented.

Layer one: Condition. The order in which the two conditions (paper-pencil and e-based) were delivered was counterbalanced. The first four students completed paper-pencil during session one and then e-based during session two. The second four students completed e-based during session one and paper-pencil during session two. This pattern repeated for the full sample. Each condition was delivered on two separate days with seven days or less between administrations.

Layer two: Maze and Slash. The order of delivering the Maze and the Slash were counterbalanced. To promote clarity for the student and reduce the need for the student to rapidly shift between different tasks, all three Maze passages were presented within the same session and all three Slash passages were presented within the same session. The student made one shift from Maze to Slash or Slash to Maze. Table 4 highlights Layer 2 of this research design: Maze and Slash. The order delivery for students 1-8 repeated for the remaining participants (i.e., 9-16, 17-24, 25-32).

Table 4

Research Design: Layer 2 (Maze and Slash)

	Order of conditions			
	Administered First		Administered Second	
1	Paper-Pencil		E-Based	
	Maze	Slash	Maze	Slash
2	Paper-Pencil		E-Based	
	Maze	Slash	Slash	Maze
3	Paper-Pencil		E-Based	
	Slash	Maze	Maze	Slash
4	Paper-Pencil		E-Based	
	Slash	Maze	Slash	Maze
5	E-Based		Paper-Pencil	
	Maze	Slash	Maze	Slash
6	E-Based		Paper-Pencil	
	Maze	Slash	Slash	Maze
7	E-Based		Paper-Pencil	
	Slash	Maze	Maze	Slash
8	E-Based		Paper-Pencil	
	Slash	Maze	Slash	Maze
…				
*n*th				

Layer three: Passage. Eight passages were used for this study, with some passages repeating and others not-repeating. Repeating probes were used to directly compare conditions. Non-repeating probes were used as a distractor to the repeating probes. These passages served to match how reading probes are delivered in the natural context, and provided a novel component in passage delivery. Within each condition (layer 1) and each measure (layer 2) the three passages (two repeating and one non-repeating) were presented to the student in a randomized order. Table 5: Passage provides an example of randomization within each cell for the students. Each reading passage was denoted by a letter (A to G).

Table 5

Research Design: Layer Three (Passages)

	Order of conditions			
	Administered First		Administered Second	
1	Paper-Pencil		E-Based	
	Maze	Slash	Maze	Slash
	ABC	FGE	DAB	EHF
2	Paper-Pencil		E-Based	
	Maze	Slash	Slash	Maze
	BAC	FGE	EHF	BDA
3	Paper-Pencil		E-Based	
	Slash	Maze	Maze	Slash
	GEF	BCA	BDA	FEH
4	Paper-Pencil		E-Based	
	Slash	Maze	Slash	Maze
	FGE	ACB	EHF	ABD
5	E-Based		Paper-Pencil	
	Maze	Slash	Maze	Slash
	BAD	FEH	ACB	FGE
6	E-Based		Paper-Pencil	
	Maze	Slash	Slash	Maze
	ABD	HEF	EGF	CAB
7	E-Based		Paper-Pencil	
	Slash	Maze	Maze	Slash
	EFH	BDA	ACB	GFE
8	E-Based		Paper-Pencil	
	Slash	Maze	Slash	Maze
	EFH	ABD	GEF	ACB
…				
*n*th				

Conditions

The independent variable was manipulated to identify if students performed reliably differently on Maze and Slash as a function of the condition (paper-pencil or e-based). Below, each condition is described with key differences outlined.

Paper-pencil condition. The Appendix provides the demonstration, practice, and passages for all paper-pencil conditions: Maze (Appendixes A to D) and Slash

(Appendixes F to J). Delivery of the paper-pencil condition generally followed a standardized administration script. The administer of the CBM tasks followed the scripted directions, provided additional training following the practice items when needed, monitored the student, recorded the time of completion, and prompted the student to move to the next passage. There was no extrinsic motivational component in the paper-pencil condition.

Out of the 40 learners, modified directions and presentation were needed for five learners. The directions were translated into American Sign Language for one student and Total Communication (English–based sign plus spoken English) for four students. The administrators followed the same testing sequence but the script was modified to match the students' communication modality. All student questions were answered in the student's preferred communication mode prior to test administration.

E-based. Avenue: PM (2011), a suite of e-based progress monitoring tools, served as the tool to deliver the e-based condition. Avenue: PM (2011) was designed to provide a set of standardized modifications specifically for use with students who are deaf and hard of hearing. Below, I describe the six modifications that differ from the paper-pencil condition (Appendixes L through S).

First, the directions were provided in a simulation format where the student could see the curser moving across the text, see the curser select the answer, and see how to change an answer when needed (Appendix L and Appendix P). Unlike the paper-pencil condition, the demonstration did not require students to listen to or watch a signed interpretation of the scripted directions. The demonstration was presented in a fully visual format, with no audio required. When needed, students re-watched or reviewed the simulation. When in the practice phase, some students also chose to repeat the practice (Appendixes M and Q).

Second, following the practice phase and after each reading probe the Avenue: PM (2011) e-based program corrected the students' performance and placed a "green check" (correct) or "red x" (incorrect) on each of the student's responses. Unlike the paper-pencil condition, students received immediate feedback after the practice item, and all reading probes. See Appendix O and Appendix S for an example of the feedback pages.

Third, Avenue: PM embedded a text movement component. For Maze, when students choose their response from the three multiple-choice options the word selected moved into the blank space within the sentence. For Slash, when the student clicked on the space between the two letters (denoting a boundary between the two words) the letters moved apart.

Fourth, the e-based condition not only allowed students to change their answers but also "erased" or removed the students' initial answer. More specifically, when the student clicked on their new answer, the old answer was automatically removed with no trace of the change. In contrast, if students changed their answer while taking the Maze in the paper-pencil condition they crossed off their initial answer and circled their preferred answer. The crossed off initial answer remained in view throughout the administration.

Fifth, the e-based condition used a visual clock that showed the passage of time as the student completes the measure (Bottom right icon in Appendix N and Appendix O). Text stating "1 minute" was also present, as a reminder of the time allotted.

Sixth, the e-based condition had a motivational feature at the bottom of the screen. If the students met a set criterion of performance the bar moved up and if their performance fell below the criterion the bar would move down. At the end of each bar was a fish icon, a smaller fish at the lower part of the bar and a bigger fish at the higher end of the bar. See bottom bar on Appendix N and Appendix R.

In summary, the independent variable was manipulated to identify if students performed reliably different in the paper-pencil condition (no modifications) or in the e-based condition (six modifications).

Procedures

As described above, procedures were followed to gain university approval, school district approval, parent consent and student assent. Participants in the study included 40 students who were deaf and hard of hearing and 21 teachers. Of the teachers, 11 selected the low involvement option (completing study forms) and 10 selected the high involvement option (delivering the CBM). Below I describe, training and administration guidelines, order of test administration, fidelity of administration, and inter-rater agreement of scoring.

Training on administration. Following the recruitment process, the 10 teachers who selected the high involvement option received training provided by me. Training included:

1. Background information regarding the purpose of the study
2. Overview and guidelines for completing study testing materials
3. Guidance on how to adhere to the testing order
4. Demonstration of all four testing conditions

One teacher had students who used Total Communication as their primary mode of communication. The teacher was provided with additional guidance on how to translate the directions for these students. Following a training session, I completed reliability checks with the teachers needing to demonstrate at least 90% accuracy prior to administration.

Order of administration. In both conditions, students were presented with demonstration and practice items prior to engaging in the testing phase. For the paper-pencil and e-based conditions, standardized administration was followed except for the one learner who used American Sign Language and the four learners who used Total Communication (English–based sign plus spoken English). The option to discontinue testing was available if the students did not understand the task; no students were discontinued.

All conditions were delivered in a one-on-one testing situation. As the visual display of Avenue: PM (2011) is the same for both a Mac or PC on a desktop or laptop, the teacher used whichever tool was available. If a laptop was selected, a teacher added a computer mouse for the student; students were not given the opportunity to use the track pad to complete the task. When the teacher did not have access to a computer, a Mac Book Air was lent to the participating teacher for use. I administered the e-based condition using a Mac Book Air.

For a student's first testing day, the teacher or I administered the first condition (paper-pencil or e-based). As each student's daily schedule was different, the second testing day occurred one to seven days after the first testing session. On the second testing day, the teacher or I administered the second condition. Depending on the student's schedule, I administered the WJ-III Passage Comprehension subtest (Woodcock et al., 2001) on either Day 1 or Day 2 of testing. If a student was absent, the testing session was completed as soon as possible.

Following all administrations, teachers who selected the high involvement option provided me with all feasibility and utility forms, all paper-pencil Maze and Slash protocols and students scores on the MAP (NWEA, 2003) spring 2015 data when available. Participating teachers who selected the high involvement option received a fifty-dollar gift card for participation after all materials were submitted.

Each teacher who selected high involvement assessed all eligible students on his or her caseload. Based on consents obtained, each teacher assessed two to three students. High involvement teachers assessed about half of the student sample 58% (*n*= 23) using the CBM tasks. I assessed the remaining sample using the CBM tasks. A trained graduate assistant or I administered the WJ-III (Woodcock et al., 2001) Passage Comprehension subtest to all participants.

Fidelity of administration. Two Fidelity of Administration checklist was generated and applied. For the CBM, there were 19 fidelity items for the paper-pencil version and 17 fidelity items for the e-based condition. Each item fell into one of four categories: preparation/introduction, Maze, Slash, and close. For the WJ-III Passage Comprehension subtest (Woodcock et al., 2001), a 17 fidelity item checklist was used. I observed teachers' CBM administrations in the paper-pencil and e-based conditions.

I also trained a research assistant who was instructed on how to use the tool and engaged in practice observations with me until 90% accuracy was obtained. The research assistant observed my CBM and WJ-III Passage Comprehension subtest (Woodcock et al., 2001) administrations. On two occasions, the research assistant observed and reported a teacher administration.

For the timed paper-pencil checklist items, a timing of administration was rated as "No" (not adequately observed) if there were three or more seconds of discrepancy between the examiner's and the observer's time documentation. Nine of the 10 teachers' full administrations (paper-pencil and e-based) were observed. Due to scheduling conflicts, only a partial administration was observed for the remaining teacher.

A portion of teacher administrations, my administrations, and research administrations were observed. For teachers, 39% (n = 9) of the student administrations were observed. For my administrations, 23% (n = 4) of CBM administrations and 21% (n = 8) of the WJ-III Passage Comprehension subtest (Woodcock, et al., 2001) administrations were observed. For the trained research assistant, 50% (n=1) of the WJ-III Passage Comprehension subtest (Woodcock, et al., 2001) administrations were observed.

Inter-rater agreement of scoring. I manually scored all paper-pencil tasks and WJ-III Passage Comprehension subtest (Woodcock et al., 2001). A trained graduate student double-scored 25% of the protocols and point-by-point inter-rater agreement (Formula: (Agree/Agree+Disagree)*100) was reported. Double scoring was not needed in the e-based condition, as it is automatically scored by the Avenue: PM suite. All data were double entered into the database and discrepancies checked. If a student's score on any measure was identified as an outlier, the score was checked again for accuracy. No errors were identified in inputting the data for these outlier scores.

Data Analysis

Five research questions were addressed in this study. Below, I describe the analyses for each research question.

Maze reliability and criterion-related validity characteristics. Research Question 1 is "Does the Maze produce scores with sufficient alternate form reliability (*r* = .80 or above) and sufficient criterion-related validity ($r \geq .50$, except IMC, r's ≥ -.50) on two administrative conditions (paper-pencil and e-based) for students who are deaf or hard of hearing?" To examine alternate form reliability of the Maze paper-pencil condition, I calculated Pearson Product Moment Correlations (*r)* for the Maze paper-pencil reading passages and used a Fisher z transformation to calculate confidence intervals for each correlation. This analysis was conducted for the four Maze scores: Correct Maze Choices (CMC), Incorrect Maze Choices (IMC), Correct Minus Incorrect Maze Choices (CMC-IMC) and Correct Minus Incorrect Maze Choices divided by 2 (CMC – IMC/2). The same procedure was used to determine the alternate form reliability for the Maze e-based condition.

To examine criterion validity of the paper-pencil Maze condition, I calculated the Pearson Product Moment Correlations (*r)* between the paper-pencil Maze reading passages and the two criterion measures: WJ-III Passage Comprehension subtest (Woodcock et al., 2001) and MAP (NWEA, 2003). I used a Fisher z transformation to generate confidence intervals for each correlation. This analysis was conducted for the three Maze scores. This same procedure was used to determine the criterion validity for the Maze e-based condition.

Slash reliability and criterion-related validity characteristics. Research Question 2 is "Does the Slash produce scores with sufficient alternate form reliability (*r*

= .80 or above) and sufficient criterion-related validity (*r* ≥ .50, except IMC, r's ≥ -.50) on two administrative conditions (paper-pencil and e-based) for students who are deaf or hard of hearing?" For this research question the four Slash metrics used included: Correct Boundaries, Incorrect Boundaries, Percent Correct and, Correct Words Identified (paper-pencil version only). Using the Slash data, the same analyses described in Research Question One were conducted for Research Question Two.

Maze: Paper-pencil vs. e-based. Research Question 3 is "Do students who are deaf or hard of hearing perform reliably differently on the Maze when it is administered in paper-pencil versus e-based conditions?" A *t*-test (two-tailed, p=.05, dependent samples) was conducted comparing student performance on the 'Race' passage delivered in the paper-pencil condition and the 'Race' passage delivered in the e-based condition for each of the four Maze scores (Bonferroni correction, p< 0.0125). A similar procedure was used for the 'Accident' passage. Additionally, collapsing across passages by taking the mean, a *t*-test (two-tailed, p =.05, dependent samples), compared student performance in paper-pencil verses e-based for the four maze scores. The Maze scores are: CMC, IMC, CMC-IMC, CMC-IMC/2.

Slash: Paper-pencil vs. e-based. Research Question 4 is: "Do students who are deaf or hard of hearing perform reliably differently on the Slash when it is administered in paper-pencil versus e-based conditions?" Using the Slash data, the same analyses described in Research Question 3 were conducted for Research Question 4. The Bonferroni correction was 0.0166. The passages "When" and "Twelve" were compared and then all passages were collapsed by taking the mean for the second analysis. The

Slash scores used for analyses were: Correct Words Identified (paper-pencil only), Correct Boundaries, Total Wrong and Percent Correct.

Feasibility of implementation and student feedback. Research Question 5 asked: “What is the feasibility of administering the paper-pencil and e-based conditions in the educational setting?” The ratings from the Usage Rating Profile – Assessment (Chafouleas et al., 2012) and Study Survey responses were summarized and reported.

Summary

This study explored whether students scores from Maze and Slash yielded sufficient reliability and criterion-related validity in paper-pencil and e-based conditions when administered to students who are deaf and hard of hearing. The effects of testing modifications delivered in an e-based format were also explored. Forty (*N*=40) students who were deaf or hard of hearing and read between the second and fifth grade reading level, participated. Students were administered CBM tasks (Maze and Slash) with performance compared to the WJ-III Passage Comprehension subtest (Woodcock, et al., 2001) and the MAP (NWEA, 2003). A correlational design was counterbalanced at the condition (paper-pencil vs. e-based) and CBM level (Maze and Slash), and randomized at the passage level. Fidelity of administration and inter-rater agreement was calculated. Data analysis included correlational analysis to explore reliability and criterion-related validity and t-tests were used to compare student performance on the e-based and paper-pencil conditions.

CHAPTER 4

RESULTS

This study explored the reliability and criterion-related validity of scores from Maze and Slash in paper-pencil and e-based conditions. The scores used for Maze included: Correct Maze Choices (CMC), Incorrect Maze Choices (IMC), Correct Maze Choices minus Incorrect Maze Choices (CMC-IMC) and Correct Maze Choices minus half the value of Incorrect Maze Choices (CMC-IMC/2). For Slash, the scores used included: Correct Words Identified, Correct Boundaries, Total Wrong (includes boundary lines that were missed and boundary lines that were incorrectly drawn), and Percent Correct (dividing the number of Correct Boundaries by the summation of Correct Boundaries and Incorrect Boundaries). All CBM scores were correlated to the criterion measures of: WJ-III Passage Comprehension subtest (Woodcock et al., 2001) and MAP (NWEA, 2003). The effect of paper-pencil versus e-based administration conditions was also compared using a *t*-test.

Research Questions

1. Does the Maze produce scores with sufficient alternate form reliability ($r \geq .80$) and sufficient criterion-related validity ($r \geq .50$, except IMC, r's $\geq$ -.50) on two administrative conditions (paper-pencil and e-based) for students who are deaf or hard of hearing?
2. Does the Slash produce scores with sufficient alternate form reliability ($r \geq .80$) and sufficient criterion-related validity ($r \geq .50$, except Total Wrong, r's $\geq$ -.80) on two administrative conditions (paper-pencil and e-based) for students who are deaf or hard of hearing?

3. Do students who are deaf or hard of hearing perform reliably differently on the Maze when it is administered in paper-pencil versus e-based conditions?
4. Do students who are deaf or hard of hearing perform reliably differently on the Slash when it is administered in paper-pencil versus e-based conditions?
5. What is the feasibility of administering the paper-pencil and e-based conditions in the educational setting?

Preliminary Analysis

Descriptive data for the full data set are presented in Table 6 for Maze and Table 7 for Slash. To determine the impact of outlier scores, I analyzed whether the data were normally distributed based on skewness and kurtosis indicators. Since non-normal distributions can inflate correlation coefficients a conservative estimate of -1 and 1 was used as the benchmark for skewness and kurtosis (Bishara & Hittner, 2015; Garson, 2012). For Maze paper-pencil, the following passages and scores fell outside the range (-1 to 1) for skewness and/or kurtosis: "Race" (IMC), "Accident" (IMC), "Pike" (CMC, IMC, CMC-IMC, & CMC-IMC/2). For Maze e-based, the following passages and scores fell outside the range for skewness and/or kurtosis: "Race" (CMC, IMC, CMC-IMC/2), "Accident" (IMC), and "Around the World" (CMC, IMC, CMC-IMC, CMC-IMC/2).

For Slash paper-pencil, the following passages and scores fell outside the range (-1 to 1) for skewness and/or kurtosis: "When" (Total Wrong, Percent Correct), "Twelve" (Correct Words Identified, Correct Boundaries), and "The" (Total Wrong, Percent Correct). For Slash e-based, the following passages and scores fell outside the range (-1 to 1) for skewness and/or kurtosis: "A" (Total Wrong, Percent Correct). Results suggest that outlier scores may be impacting the normality of the distributions for Maze and Slash.

To explore the impact of outliers on the data distribution, I identified outliers for each of the 45 scores generated from the CBM. Subsequently I counted and recorded how many scores were identified as an outlier score for each participant. Of the 40 participants, 24 students did not have any outlier scores, 13 had one to three outlier scores, 1 student had 9 outlier scores and 2 students had 10 outlier scores. Based on this analysis, the three participants who had the highest frequency of outlier scores were dropped from the analysis (one student below the group, two students above the group). Tables 8 and 9 present the reduced dataset that is less three participants (n = 37).

Table 6

Maze: Descriptive Data for Full Dataset

	N	Mean	*SD*	Range	Skewness	Kurtosis
Paper-Pencil						
Race						
Correct Maze Choices (CMC)	40	8.10	3.98	2.00-18.62	.86	.40
Incorrect Maze Choices (IMC)	40	.48	.96	0.00-5.00	3.03	10.58
CMC-IMC	40	7.62	4.30	1.00-18.62	.74	.11
CMC-IMC/2	40	7.86	4.12	1.50-18.62	.83	.28
Accident						
Correct Maze Choices (CMC)	40	6.81	4.17	1.00-18.31	.86	.01
Incorrect Maze Choices (IMC)	40	.57	.84	0.00-3.00	1.14	.07
CMC-IMC	40	6.23	4.71	-2.00-18.31	.58	-.32
CMC-IMC/2	40	6.52	4.43	-.50-18.31	.73	-0.17
Pike						
Correct Maze Choices (CMC)	40	7.34	3.77	2.00-18.00	1.06	.76
Incorrect Maze Choices (IMC)	40	1.08	3.06	0.00-19.20	5.16	27.62
CMC-IMC	40	6.26	5.38	-15.60-18.00	-.14	5.22
CMC-IMC/2	40	6.80	4.38	-6.00-18.00	.22	1.35
E-Based						
Race						
Correct Maze Choices (CMC)	40	8.55	3.58	4.00-18.62	1.2	1.02
Incorrect Maze Choices (IMC)	40	1.05	1.48	0.00-8.00	2.67	9.44
CMC-IMC	40	7.50	4.30	0.00-18.62	.73	.19
CMC-IMC/2	40	8.02	3.89	3.50-18.62	1.04	.59
Accident						
Correct Maze Choices (CMC)	40	7.60	3.64	2.00-16.00	.31	-0.70
Incorrect Maze Choices (IMC)	40	.92	2.09	0.00-13.00	4.81	24.78
CMC-IMC	40	6.67	4.62	-8.00-16.00	-.61	0.84
CMC-IMC/2	40	7.14	4.02	-1.50-16.00	0.02	-.65
Around the World						
Correct Maze Choices (CMC)	40	8.83	4.43	2.00-24.55	1.52	2.65
Incorrect Maze Choices (IMC)	40	.48	.68	0.00-2.00	1.05	-.21
CMC-IMC	40	8.35	4.66	1.00-24.55	1.39	2.39
CMC-IMC/2	40	8.59	4.54	1.50-24.55	1.47	2.56

Table 7

Slash: Descriptive Data for Full Dataset

	N	Mean	*SD*	Range	Skewness	Kurtosis
Paper-Pencil						
When						
Correct Words Identified	40	30.42	12.15	7.00-55.38	.28	-0.65
Correct Boundaries	40	23.60	8.66	6.00-41.54	.27	-.55
Total Wrong	40	2.50	3.45	0.00-13.00	1.43	1.36
Percent Correct	40	88.89	15.74	31.58-100.00	-1.65	2.57
Twelve						
Correct Words Identified	40	25.00	10.67	8.00-46.29	.35	-1.13
Correct Boundaries	40	19.49	7.53	7.00-34.29	.32	-1.05
Total Wrong	40	5.00	4.88	0.00-16.00	.68	-.71
Percent Correct	40	78.66	20.72	30.43-100.00	-.57	-.98
The						
Correct Words Identified	40	31.30	15.35	8.57-66.67	.40	-.68
Correct Boundaries	40	26.08	11.85	8.00-53.33	.37	-.69
Total Wrong	40	3.00	4.42	0.00-18.00	1.60	2.04
Percent Correct	40	87.72	17.27	42.11-100.00	-1.21	.12
E-Based						
When						
Correct Boundaries	40	16.67	7.35	2.00-32.73	.16	-.05
Total Wrong	40	4.87	5.44	0.00-19.00	.89	-.29
Percent Correct	40	75.92	26.36	9.52-100.00	-.85	-.29
Twelve						
Correct Boundaries	40	16.30	6.24	4.00-28.57	0.16	-.71
Total Wrong	40	5.80	5.22	0.00-18.00	.55	-.87
Percent Correct	40	73.16	23.92	20.00-100.00	-0.50	-.99
A						
Correct Boundaries	40	18.48	6.87	6.00-39.13	.62	.43
Total Wrong	40	3.23	2.86	0.00-12.27	1.21	1.21
Percent Correct	40	83.39	15.57	37.50-100.00	-1.17	.61

Table 8

Maze: Descriptive Data for Reduced Dataset

	N	Mean	*SD*	Range	Skewness	Kurtosis
Paper-Pencil						
Race						
Correct Maze Choices (CMC)	37	7.57	3.30	2.00-16.00	.43	-.37
Incorrect Maze Choices (IMC)	37	0.38	.64	0.00-3.00	2.00	5.00
CMC-IMC	37	7.19	3.53	1.00-16.00	.35	-.48
CMC-IMC/2	37	7.38	3.40	1.50-16.00	.40	-.41
Accident						
Correct Maze Choices (CMC)	37	6.43	3.50	1.00-15.00	.72	-.35
Incorrect Maze Choices (IMC)	37	.54	.77	0.00-2.00	.95	-.70
CMC-IMC	37	5.89	3.98	-1.00-15.00	.49	-.61
CMC-IMC/2	37	6.16	3.73	0.00-15.00	.61	-.49
Pike						
Correct Maze Choices (CMC)	37	6.89	3.01	2.00-15.00	.66	.09
Incorrect Maze Choices (IMC)	37	.65	.89	0.00-3.00	1.18	.39
CMC-IMC	37	6.24	3.32	-1.00-15.00	.38	.20
CMC-IMC/2	37	6.57	3.14	0.50-15.00	.54	.16
E-Based						
Race						
Correct Maze Choices (CMC)	37	8.03	2.85	4.00-16.00	.84	.03
Incorrect Maze Choices (IMC)	37	.92	.98	0.00-3.00	.84	-.36
CMC-IMC	37	7.11	3.41	2.00-15.00	.41	-.57
CMC-IMC/2	37	7.57	3.11	3.50-15.50	.63	-.21
Accident						
Correct Maze Choices (CMC)	37	7.24	3.25	2.00-14.00	.02	-1.09
Incorrect Maze Choices (IMC)	37	.62	.76	0.00-3.00	1.46	2.48
CMC-IMC	37	6.62	3.63	-1.00-13.00	-.21	-1.01
CMC-IMC/2	37	6.93	3.43	0.50-13.50	-.09	-1.10
Around the World						
Correct Maze Choices (CMC)	37	8.22	3.22	2.00-17.00	.73	.47
Incorrect Maze Choices (IMC)	37	.46	.65	0.00-2.00	1.03	-.14
CMC-IMC	37	7.76	3.40	1.00-17.00	.60	.49
CMC-IMC/2	37	7.99	3.30	1.5-17.00	.68	.51

Table 9

Slash: Descriptive Data for Reduced Dataset

	N	Mean	SD	Range	Skewness	Kurtosis
Paper-Pencil						
When						
Correct Words Identified	37	29.54	11.87	7.00-55.38	.29	-.69
Correct Boundaries	37	22.92	8.43	6.00-41.54	.29	-.55
Total Wrong	37	2.37	3.14	0.00-13.00	1.38	1.48
Percent Correct	37	88.81	15.85	31.58-100.00	-1.70	2.75
Twelve						
Correct Words Identified	37	24.26	10.12	8.00-44.12	.34	-1.08
Correct Boundaries	37	18.94	7.20	7.00-33.53	.32	-1.00
Total Wrong	37	4.99	4.63	0.00-16.00	.63	-.69
Percent Correct	37	78.28	20.41	30.43-100.00	-.58	-.89
The						
Correct Words Identified	37	30.48	15.62	8.57-66.67	.53	-.61
Correct Boundaries	37	25.27	11.92	8.00-53.33	.53	-.51
Total Wrong	37	2.70	3.81	0.00-13.47	1.30	.71
Percent Correct	37	87.74	17.53	42.11-100.00	-1.22	.13
E-Based						
When						
Correct Boundaries	37	16.66	7.56	2.00-32.73	.15	-.59
Total Wrong	37	4.87	5.55	0.00-19.00	.90	-.33
Percent Correct	37	75.77	27.02	9.52-100.00	-.85	-.38
Twelve						
Correct Boundaries	37	16.13	6.40	4.00-28.57	.22	-.76
Total Wrong	37	5.99	5.26	0.00-18.00	.52	-.91
Percent Correct	37	72.17	24.11	20.00-100.00	-.46	-1.04
A						
Correct Boundaries	37	18.60	7.02	6.00-39.13	.59	.36
Total Wrong	37	3.08	2.51	0.00-10.00	.92	.29
Percent Correct	37	83.80	15.11	37.50-100.00	-1.27	1.02

Table 10 displays the distribution statistics for the criterion measures. For the WJ-III: Passage Comprehension subtest (Woodcock et al., 2001), the full data set included all students and the modified data set was all students, less the three students who were dropped. As the MAP (NWEA, 2003) was not required by all districts, not all students participants had a MAP score (27 of 40 participants). Two of the dropped students had taken the MAP, and thus the full dataset was reduced by two for the MAP (NWEA, 2003).

Table 10

Criterion Measures: Full and Reduced Dataset

	N	*M*	*SD*	Range	Skewness	Kurtosis
WJ-III Passage Comprehension						
Full Dataset	40	26.07	4.65	14.00-35.00	-0.47	-0.04
Reduced Dataset	37	26.08	4.17	17.00-35.00	-0.23	-0.20
MAP						
Full Dataset	27	205.07	12.58	180.00-227.00	-0.11	-1.02
Reduced Dataset	25	203.64	11.91	180.00-227.00	-0.09	-0.94

Note: WJ-III Passage Comprehension *M* = raw score; MAP *M* = RIT score.

The removal of the three students who had the highest frequency of outlier scores generally resulted in a more acceptable distribution of score types. This adjustment yielded the following score types to move into the more acceptable ranges (within -1 and 1 for skewness and kurtosis): Paper-Pencil: “Accident” (IMC), “Pike” (CMC, CMC-IMC, CMC-IMC/2); e-based: “Race” (CMC, IMC, CMC-IMC/2), and “Around the World” (CMC, CMC-IMC, CMC-IMC/2). With this adjustment, three scores types (e-based, “Accident” CMC, CMC-IMC, CMC-IMC/2) moved to falling just outside the range (-1 to 1). Three score types continued to fall outside the range, however, the skewness and kurtosis values moved closer to the acceptable ranges (Paper-Pencil: “Race” IMC, “Pike” IMC, e-based: “Accident” IMC, “Around the World” IMC).

For Slash, the movement towards greater normality of the distribution of score types was less notable. Seven of the score types continued to have skewness and kurtosis scores outside of -1 and 1. This adjustment caused one score type to move outside the skeweness or kurtosis range (e-based: “Twelve,” Percent Correct) and one score type moved within the range (e-based “A,” Total Wrong). Globally, this adjustment yielded score types that were more normally distributed with the most notable movement toward normality for Maze score types.

Maze Reliability and Criterion-Related Validity Characteristics

Research Question 1 examined the alternate form and criterion-related validity of Maze scores. Inter-rater agreement was 100% for the paper-pencil Maze data (10 of 40 participant cases were reviewed, 25% of the dataset). Inter-rater agreement was not required for the e-based condition due to the automatic scoring.

Reliability. Pearson Product Moment correlations for the passages delivered in the paper-pencil condition and the e-based condition are displayed in Table 11. For the paper-pencil condition, the “Race”-“Accident” correlations were sufficient ($r \geq .80$) and significant ($p<.01$) for the scores of: CMC ($r = .84$, $p< .01$), CMC-IMC ($r=.84$, $p< .01$), and CMC-IMC/2 ($r = .84$, $p <.01$). For the e-based condition, the “Race”- “Around the World” correlations were sufficient ($r \geq .80$) and significant ($p<.01$) for the correlation of CMC ($r = .80$, $p<.01$).

Table 11

Maze: Alternate-Form Reliability Coefficients

	N	*r*	CI (95%) LL	UL	*t* (35)	*p*
Paper-Pencil						
Race-Accident						
Correct Maze Choices (CMC)	37	**0.84**	.72	.92	9.35	< .01
Incorrect Maze Choices (IMC)	37	0.48	.18	.69	3.21	< .01
CMC-IMC	37	**0.84**	.70	.91	9.00	< .01
CMC-IMC/2	37	**0.84**	.71	.92	9.27	< .01
Race-Pike						
Correct Maze Choices (CMC)	37	0.76	.57	.87	6.86	< .01
Incorrect Maze Choices (IMC)	37	0.04	-.28	.36	0.27	.79
CMC-IMC	37	0.66	.43	.81	5.20	< .01
CMC-IMC/2	37	0.71	.51	.84	6.04	< .01
Pike-Accident						
Correct Maze Choices (CMC)	37	0.78	.61	.88	7.30	< .01
Incorrect Maze Choices (IMC)	37	0.12	-.21	.43	0.74	.47
CMC-IMC	37	0.73	.53	.85	6.25	< .01
CMC-IMC/2	37	0.76	.58	.87	6.89	< .01
E-based						
Race-Accident						
Correct Maze Choices (CMC)	37	0.74	.55	.86	6.5	< .01
Incorrect Maze Choices (IMC)	37	0.52	.23	.72	3.57	< .01
CMC-IMC	37	0.72	.52	.85	6.15	< .01
CMC-IMC/2	37	0.73	.54	.85	6.39	< .01
Race-Around the World						
Correct Maze Choices (CMC)	37	**0.80**	.64	.89	7.80	< .01
Incorrect Maze Choices (IMC)	37	-0.03	-.35	.30	-0.16	.87
CMC-IMC	37	0.66	.43	.81	5.26	< .01
CMC-IMC/2	37	0.73	.54	.86	6.40	< .01
Around the World-Accident						
Correct Maze Choices (CMC)	37	0.66	.43	.81	5.27	< .01
Incorrect Maze Choices (IMC)	37	0.31	-.02	.57	1.90	< .01
CMC-IMC	37	0.61	.35	.78	4.51	< .01
CMC-IMC/2	37	0.64	.39	.80	4.88	< .01

Note. Coefficients ≥.80 in boldface. *LL* = lower limited; *UL* = upper limit. CMC-IMC = Correct Maze Choices – Incorrect Maze Choices; CMC-IMC/2= Correct Maze Choices – (Incorrect Maze Choices/2)

Criterion-related validity. Pearson Product Moment correlations for the passages delivered in the paper-pencil condition and the e-based condition for Maze are displayed in Table 12 and 13. For the paper-pencil condition, the WJ-III Passage comprehension subtest (Woodcock et al., 2001) correlations with the Maze passages were sufficient for "Race" (CMC-IMC) and "Accident" (CMC, CMC-IMC, CMC-IMC/2). For MAP (NWEA, 2003), correlations with the Maze passages were sufficient for the following passages and scores: "Race" (CMC, CMC-IMC, CMC-IMC/2), "Accident" (CMC, IMC, CMC-IMC, CMC-IMC/2), and Pike (CMC-IMC, CMC-IMC/2). Each of these correlations was significant (*ps* $\leq$ .01).

For the e-based condition, the WJ-III Passage comprehension subtest (Woodcock et al., 2001) correlations with Maze passages were sufficient for the following passages and scores: "Race" (CMC, CMC-IMC, CMC-IMC/2), "Accident" (IMC), and "Around the World" (CMC-IMC, CMC-IMC/2). For the MAP (NWEA, 2003), correlations with the Maze passages were sufficient for the following passages and scores: "Race"(CMC, CMC-IMC, CMC-IMC/2), "Accident" (CMC, CMC-IMC, CMC-IMC/2), and "Around the World" (CMC, CMC-IMC, CMC-IMC/2). Each of these correlations were significant (*ps* $\leq$ .01).

Table 12

Maze: Criterion-Related Validity Coefficients for the Paper-Pencil Condition

	N	r	CI (95%) LL	UL	*t*	*p*
Paper-Pencil						
Race – WJ-III Passage Comprehension						
Correct Maze Choices (CMC)	37	.48	.18	.70	3.24	<.01
Incorrect Maze Choices (IMC)	37	-.27	-.55	.06	-1.68	.10
CMC-IMC	37	**.50**	.21	.71	3.40	<.01
CMC-IMC/2	37	.49	.20	.70	3.34	<.01
Race-MAP						
Correct Maze Choices (CMC)	25	**0.61**	.28	.81	3.66	<.01
Incorrect Maze Choices (IMC)	25	-0.29	-.62	.11	-1.47	.15
CMC-IMC	25	**0.60**	.27	.81	3.64	<.01
CMC-IMC/2	25	**0.61**	.28	.81	3.65	<.01
Accident- WJ-III Passage Comprehension						
Correct Maze Choices (CMC)	37	**.57**	.30	.76	4.11	<.01
Incorrect Maze Choices (IMC)	37	-.43	-.66	-.12	-2.83	<.01
CMC-IMC	37	**.58**	.32	.76	4.26	<.01
CMC-IMC/2	37	**.58**	.32	.76	4.21	<.01
Accident - MAP						
Correct Maze Choices (CMC)	25	**.65**	.35	.83	4.12	<.01
Incorrect Maze Choices (IMC)	25	**-.50**	-.75	-.13	-2.74	.01
CMC-IMC	25	**.67**	.38	.84	4.36	<.01
CMC-IMC/2	25	**.67**	.37	.84	4.27	<.01
Pike - WJ-III Passage Comprehension						
Correct Maze Choices (CMC)	37	.33	.01	.59	2.06	.05
Incorrect Maze Choices (IMC)	37	-.32	-.59	.00	-2.01	.05
CMC-IMC	37	.38	.07	.63	2.46	.02
CMC-IMC/2	37	.36	.04	.61	2.29	.03
Pike - MAP						
Correct Maze Choices (CMC)	25	.49	.12	.74	2.69	.01
Incorrect Maze Choices (IMC)	25	-.20	-.55	.21	-1.00	.33
CMC-IMC	25	**.50**	.13	.75	2.74	.01
CMC-IMC/2	25	**.50**	.13	.75	2.75	.01

Note. Coefficients ≥±.50 in boldface. *LL* = lower limited; *UL* = upper limit. CMC-IMC = Correct Maze Choices – Incorrect Maze Choices; CMC-IMC/2= Correct Maze Choices – (Incorrect Maze Choices/2); ; WJ-III Passage Comprehension = Woodcock Johnson: Test of Achievement: Third Edition: Passage comprehension subtest

Table 13

Maze: Criterion-Related Validity Coefficients the E-Based Condition

	N	r	CI (95%) LL	CI (95%) UL	*t*	*p*
E-based						
Race – WJ-III Passage Comprehension						
Correct Maze Choices (CMC)	37	**.50**	.21	.71	3.42	<.01
Incorrect Maze Choices (IMC)	37	-.33	-.59	-.01	-2.07	.05
CMC-IMC	37	**.51**	.23	.72	3.54	<.01
CMC-IMC/2	37	**.51**	.22	.72	3.52	<.01
Race-MAP						
Correct Maze Choices (CMC)	25	**.57**	.22	.79	3.31	<.01
Incorrect Maze Choices (IMC)	25	-.04	-.43	.36	-.20	.84
CMC-IMC	25	**.51**	.14	.75	2.83	<.01
CMC-IMC/2	25	**.54**	.18	.77	3.08	.01
Accident- WJ-III Passage Comprehension						
Correct Maze Choices (CMC)	37	.43	.12	.66	2.83	<.01
Incorrect Maze Choices (IMC)	37	**-.50**	-.71	-.21	-3.42	<.01
CMC-IMC	37	.49	.20	.70	3.33	<.01
CMC-IMC/2	37	.46	.17	.69	3.10	<.01
Accident - MAP						
Correct Maze Choices (CMC)	25	**.56**	.21	.78	3.21	<.01
Incorrect Maze Choices (IMC)	25	-.16	-.52	.25	-.78	.44
CMC-IMC	25	**.54**	.18	.77	3.01	<.01
CMC-IMC/2	25	**.55**	.20	.78	3.15	<.01
Around the World - WJ-III Passage Comprehension						
Correct Maze Choices (CMC)	37	.48	.18	.69	3.21	<.01
Incorrect Maze Choices (IMC)	37	-.37	-.62	-.06	-2.38	.02
CMC-IMC	37	**.52**	.24	.72	3.63	<.01
CMC-IMC/2	37	**.50**	.21	.71	3.43	<.01
Around the World - MAP						
Correct Maze Choices (CMC)	25	**.60**	.27	.80	3.58	<.01
Incorrect Maze Choices (IMC)	25	-.18	-.54	.23	-0.87	.40
CMC-IMC	25	**.61**	.28	.81	3.67	<.01
CMC-IMC/2	25	**.61**	.28	.81	3.65	<.01

Note. Coefficients ≥±.50 in boldface. *LL* = lower limited; *UL* = upper limit. CMC-IMC = Correct Maze Choices – Incorrect Maze Choices; CMC-IMC/2= Correct Maze Choices – (Incorrect Maze Choices/2); WJ-III Passage Comprehension = Woodcock Johnson: Test of Achievement: Third Edition: Passage comprehension subtest

Slash Reliability and Criterion-Related Validity Characteristics

Research Question 2 examined the alternate form and criterion-related validity of Slash scores. Inter-rater agreement was 95% to 100% (*M* = 100%) for the paper-pencil condition (10 of 40 participant cases were reviewed, 25% of the data set). Of the 21 scores reviewed for each participant, there was one disagreement in one participant's file that was resolved. Inter-rater agreement was not required for the e-based condition due to the automatic scoring.

Reliability. Pearson Product Moment correlations for the passages delivered in the paper-pencil condition and the e-based condition are displayed in Table 14. For the paper-pencil condition, all correlations fell outside of the sufficient range.

Table 14

Slash: Alternate-Form Reliability Coefficients

	N	r	CI (95%)		*t* (35)	*p*
			LL	UL		
Paper-Pencil						
Twelve-When						
Correct Words Identified	37	.69	.47	.83	5.60	<.01
Correct Boundaries	37	.70	.49	.84	5.85	<.01
Total Wrong	37	.57	.30	.75	4.07	<.01
Percent Correct	37	.59	.33	.77	4.34	<.01
When-The						
Correct Words Identified	37	.64	.40	.80	4.95	<.01
Correct Boundaries	37	.64	.40	.80	4.96	<.01
Total Wrong	37	.46	.16	.68	3.03	<.01
Percent Correct	37	.46	.16	.68	3.04	<.01
The-Twelve						
Correct Words Identified	37	.50	.22	.71	3.46	<.01
Correct Boundaries	37	.51	.22	.71	3.47	<.01
Total Wrong	37	.42	.11	.65	2.70	.01
Percent Correct	37	.39	.07	.63	2.49	.02
E-based						
Twelve-When						
Correct Boundaries	37	.75	.56	.86	6.68	<.01
Total Wrong	37	.69	.47	.83	5.63	<.01
Percent Correct	37	.71	.49	.84	5.89	<.01
When -A						
Correct Boundaries	37	.68	.45	.82	5.45	<.01
Total Wrong	37	.45	.15	.68	3.02	<.01
Percent Correct	37	.49	.20	.71	3.36	<.01
A- Twelve						
Correct Boundaries	37	.73	.52	.85	6.23	<.01
Total Wrong	37	.41	.10	.65	2.69	.01
Percent Correct	37	.53	.24	.73	3.65	<.01

Note: Coefficients ≥.80 in boldface.

Criterion-related validity. Pearson Product Moment correlations for the passages delivered in the paper-pencil condition and the e-based condition for Slash are displayed in Table 15 and 16. For the paper-pencil condition, the WJ-III Passage comprehension subtest (Woodcock et al., 2001) correlation with the "Twelve" passage

(Total Wrong score) was sufficient and significant (r =-.51, p<.01). For the MAP (NWEA, 2003), correlations with the Slash were sufficient for the following passages and scores: “When” (Correct Words Identified, Correct Boundaries, Total Wrong, and Percent Correct), “Twelve” (Correct Words Identified, Correct Boundaries, Total Wrong, and Percent Correct), and “The”(Correct Words Identified Correct Boundaries). Each of these correlations were significant (ps < .01).

For the e-based condition, the WJ-III Passage comprehension subtest (Woodcock et al., 2001) correlation with “Twelve” (Total Wrong, Percent Correct) was sufficient and significant (p<.01). For the MAP (NWEA, 2003), correlations with the Slash were sufficient and significant (ps <.01) for the “Twelve” passasge (Correct Boundaries, Total Wrong, Percent Correct).

Table 15

Slash: Criterion-Related Validity Coefficients for the Paper-Pencil Condition

	N	r	CI (95%) LL	UL	*t* (35)	*p*
Paper-Pencil						
When- WJ-III Passage Comprehension						
Correct Words Identified	37	.40	.09	.64	2.60	.01
Correct Boundaries	37	.41	.10	.65	2.67	.01
Total Wrong	37	-.26	-.66	-.12	-1.57	.12
Percent Correct	37	.27	-.06	.55	1.67	.10
When-MAP						
Correct Words Identified	25	**.57**	.23	.79	3.35	<.01
Correct Boundaries	25	**.55**	.19	.77	3.12	<.01
Total Wrong	25	**-.54**	-.77	-.19	-3.10	<.01
Percent Correct	25	**.57**	.22	.79	3.32	<.01
Twelve - WJ-III Passage Comprehension						
Correct Words Identified	37	.48	.19	.70	3.27	<.01
Correct Boundaries	37	.45	.15	.68	3.02	<.01
Total Wrong	37	**-.51**	-.66	-.12	-3.5	<.01
Percent Correct	37	.49	.20	.70	3.36	<.01
Twelve-MAP						
Correct Words Identified	25	**.71**	.44	.86	4.84	<.01
Correct Boundaries	25	**.72**	.45	.87	4.93	<.01
Total Wrong	25	**-.71**	-.86	-.44	-4.82	<.01
Percent Correct	25	**.73**	.46	.87	5.06	<.01
The - WJ-III Passage Comprehension						
Correct Words Identified	37	.33	.01	.59	2.06	.05
Correct Boundaries	37	.30	-.03	.57	1.86	.07
Total Wrong	37	-.38	-.63	-.07	-2.45	.02
Percent Correct	37	.40	.09	.64	2.58	.01
The-MAP						
Correct Words Identified	25	**.55**	.19	.77	3.12	<.01
Correct Boundaries	25	**.53**	.17	.77	3.00	<.01
Total Wrong	25	-.42	-.70	-.03	-2.20	.04
Percent Correct	25	.44	.05	.71	2.33	.03

Note: Coefficients ≥±.50 in boldface. WJ-III Passage Comprehension = Woodcock Johnson: Test of Achievement: Third Edition: Passage comprehension subtest

Table 16

Slash: Criterion-Related Validity Coefficients for the E-Based Condition

	N	r	CI (95%) LL	UL	*t*	*p*
E-based						
When- WJ-III Passage Comprehension						
Correct Boundaries	37	.27	-.06	.55	1.68	.10
Total Wrong	37	-.23	-.51	.10	-1.38	.17
Percent Correct	37	.26	-.07	.54	1.58	.12
When-MAP						
Correct Boundaries	25	.48	.10	.74	2.62	.02
Total Wrong	25	-.46	-.72	-.08	-2.47	.02
Percent Correct	25	.47	.09	.73	2.53	.02
Twelve - WJ-III Passage Comprehension						
Correct Boundaries	37	.34	.02	.60	2.12	.04
Total Wrong	37	**-.53**	-.73	-.25	-3.68	<.01
Percent Correct	37	**.50**	.21	.71	3.44	<.01
Twelve-MAP						
Correct Boundaries	25	**.51**	.14	.75	2.82	<.01
Total Wrong	25	**-.59**	-.80	-.25	-3.49	<.01
Percent Correct	25	**.58**	.24	.79	3.43	<.01
A - WJ-III Passage Comprehension						
Correct Boundaries	37	.25	-.09	.53	1.50	.14
Total Wrong	37	.05	-.28	.37	0.30	.76
Percent Correct	37	.08	-.25	.40	0.50	.62
A-MAP						
Correct Boundaries	25	.37	-.03	.67	1.90	.07
Total Wrong	25	-.21	-.56	.20	-1.02	.32
Percent Correct	25	.28	-.13	.61	1.39	.18

Note: Coefficients ≥±.50 in boldface. WJ-III Passage Comprehension = Woodcock Johnson: Test of Achievement: Third Edition: Passage comprehension subtest

Maze: Paper-pencil vs. e-based. Research question three examined the type of administration on student performance on the Maze. The data were analyzed by comparing student performance for each passage in both conditions and across passages in both conditions.

Comparison of individual reading passages. Two passages, 'Race' and 'Accident', were delivered in the paper-pencil and e-based conditions. A paired sample *t*-test (two-tailed, p =.05) was used to compare the average performance across students on the 'Race' delivered in the paper-pencil condition to the average performance across students on the 'Race' delivered in the e-based condition (Table 17). This comparison was completed for the four Maze 'Race' scores (CMC, IMC, CMC-IMC, CMC-IMC/2). As there were four variables for comparison in the "Race" dataset, a Bonferroni correction was used. The Bonferroni correction (α/n = .05/4) yielded a p value of 0.0125. For the "Race" passage, the IMC score was statistically significant ($p < .01$), favoring the paper-pencil condition. All other scores were non-significant.

This same process was used to compare student performance on the "Accident" passage in the paper-pencil condition to student performance on the "Accident" passage the e-based condition, for each of the four scores (CMC, IMC, CMC-IMC, and CMC-IMC/2). The same Bonferroni correction was used. No comparisons were statistically significant at the p < is 0.0125 level for the "Accident" passages.

Table 17

Paired Sample T-Test Comparing Paper-Pencil Student Performance Scores to Computer Student Performance Scores for Each Maze Passage and Each Score

	Paper-Pencil		Computer		*t* (36)	*p*	95% CI	
	M	*SD*	*M*	*SD*			LL	UL
Race								
CMC	7.57	3.30	8.03	2.85	-1.11	0.28	-1.30	0.38
IMC	0.38	0.64	0.92	0.98	-3.33	**<.01**	-0.87	-0.21
CMC-IMC	7.19	3.53	7.11	3.41	0.16	0.87	-0.92	1.08
CMC-IMC/2	7.38	3.40	7.57	3.11	-0.42	0.68	-1.10	0.72
Accident								
CMC	6.43	3.50	7.24	3.25	-1.76	0.09	-1.74	0.12
IMC	0.54	0.77	0.62	0.76	-0.65	0.52	-0.33	0.17
CMC-IMC	5.89	3.98	6.62	3.63	-1.41	0.17	-1.78	0.32
CMC-IMC/2	6.16	3.73	6.93	3.43	-1.58	0.12	-1.76	0.22

Note. $p < 0.01$ in boldface (p value after Bonferroni correction). CI= confidence interval; LL = lower limit; UL = Upper limit; CMC= Correct Maze Choices; IMC = Incorrect Maze Choices; CMC-IMC = Correct Maze Choices – Incorrect Maze Choices; CMC-IMC/2= Correct Maze Choices – (Incorrect Maze Choices/2).

Mean performance across probes. To provide a more stable estimate of student performance, the mean of student performance across all passages were calculated. For example, Student A's CMC score on the "Accident," "Race," and "Pike" passages were averaged. Then, each student's new averaged score was averaged across participants to gain a mean score for CMC paper-pencil condition. This same calculation was completed for the CMC score on the "Accident," "Race," and "Around the World" for the e-based condition. A paired sample *t*-test (two-tailed, p =.05) was used to compare these two mean values (Table 18). This same process was completed for the other Maze scores as well (CMC, IMC, CMC-IMC, and CMC-IMC/2). As there were multiple comparisons on the same dataset, a Bonferroni correction was used. The Bonferroni correction (α/n =

.05/4) yielded a *p* value of 0.0125. There was a statistically significant difference for CMC ($p < .01$) between the paper-pencil and e-based conditions, favoring the e-based condition. All other comparisons were non-significant.

Table 18

Paired Sample T-Test Comparing Mean Paper-Pencil Student Performance Scores to Mean E-based Student Performance Scores Across Scores for Maze

	Paper-Pencil		Computer		*t* (36)	*p*	95% CI	
	M	*SD*	*M*	*SD*			LL	UL
CMC	6.96	3.04	7.83	2.82	-2.71	**0.01**	-1.51	-0.22
IMC	0.52	0.52	0.67	0.58	-2.01	0.05	-0.29	0.00
CMC-IMC	6.44	3.29	7.16	3.07	-2.19	0.03	-1.39	-0.05
CMC-IMC/2	6.70	3.16	7.50	2.93	-2.47	0.02	-1.45	-0.14

Note. $p < 0.01$ in boldface (*p* value after Bonferroni correction). CI= confidence interval; LL = lower limit; UL = Upper limit; CMC= Correct Maze Choices; IMC = Incorrect Maze Choices; CMC-IMC = Correct Maze Choices – Incorrect Maze Choices; CMC-IMC/2= Correct Maze Choices – (Incorrect Maze Choices/2).

Slash: Paper-pencil vs. e-based. Research question four examined the influence of the independent variable (administration type: paper-pencil vs. e-based) on student performance on the Slash. The data were analyzed by comparing student performance for each passage in both conditions and across passages in both conditions.

Comparison of individual reading passages. Two passages, “When” and “Twelve”, were delivered in the paper-pencil and e-based conditions. A paired sample *t*-test (two-tailed, $p = .05$) was used to compare the average performance across students on the “When” passage delivered in the paper-pencil condition to the average performance across students on the “When” delivered in the e-based condition. This comparison was completed for the three Slash “Twelve” scores (Correct Boundaries, Total Wrong and Percent Correct) (Table 19). As the Correct Words Identified is not present in the e-based condition, this score was not used in comparison. As there were three variables for comparison in the "When" dataset, a Bonferroni correction was used. The Bonferroni correction (α/n = .05/3) yielded a *p* value of 0.0166. All three scores for the "When" passage were statistically significant favoring the paper-pencil condition, below $p <$ 0.0166 (all *p*'s < .01).

This same process was used to compare student performance on the "Twelve" passage in the paper-pencil condition to student performance on the "Twelve" passage the e-based condition, for each of the four scores (Correct Boundaries, Total Wrong and Percent Correct). The same bonferroni correction was used. No comparisons were statistically significant at the $p <$ is 0.0166 level for the "Twelve" passages.

Table 19

Paired Sample T-test Comparing Paper-Pencil Student Performance Scores to E-Based Student Performance Scores for Each Slash Passage and Each Score

	Paper-Pencil		Computer		*t* (36)	*p*	95% CI	
	M	*SD*	*M*	*SD*			LL	UL
When								
Correct Bound.	22.92	8.43	16.66	7.56	4.93	**<0.01**	3.68	8.83
Total Wrong	2.37	3.14	4.87	5.55	-3.74	**<0.01**	-3.87	-1.15
Percent Correct	88.81	15.85	75.77	27.02	3.87	**<0.01**	6.20	19.88
Twelve								
Correct Bound.	18.94	7.2	16.13	6.4	2.31	0.03	0.35	5.27
Total Wrong	4.99	4.63	5.99	5.26	-1.70	0.10	-2.21	0.20
Percent Correct	78.28	20.41	72.17	24.11	2.19	0.04	0.45	11.76

Note: $p < 0.02$ in boldface (*p* value after Bonferroni correction). Correct Bound = Correct Boundaries

Mean performance across probes. To provide a more stable estimate of student performance, the mean of student performance across all Slash passages were calculated.

The same process described for the Maze was used for the Slash. A paired sample *t*-test (two-tailed, *p* =.05) was used to compare the mean of the paper-pencil Slash condition to the e-based Slash condition for three scores (Correct Boundaries, Total Wrong, Percent Correct) (Table 20). As there were multiple comparisons on the same dataset, a bonferroni correction was used. The bonferroni correction (1/*n* = .05/3) yielded a *p* value of 0.0166. There was a statistically significant difference, favoring the paper-pencil condition, for all three Slash scores (all *p*'s < .01).

Table 20

Paired Sample T-Test Comparing Slash Paper-Pencil Student Performance Scores to Mean E-Based Student Performance Scores Across Four Maze Scores

	Paper-Pencil		Computer		*t* (36)	*p*	CI 95%	
	M	*SD*	*M*	*SD*			LL	UL
Correct Boundaries	22.38	7.93	17.13	6.30	4.69	**<0.01**	2.98	7.52
Total Wrong	3.35	3.13	4.65	3.78	-3.08	**<0.01**	-2.15	-0.44
Percent Correct	84.94	14.49	77.25	18.99	3.44	**<0.01**	3.16	12.23

Note. *p* < 0.02 in boldface (*p* value after Bonferroni correction). CI= confidence interval; LL = lower limit; UL = Upper limit

Feasibility of Implementation and Student Feedback

Research question five explored teachers' perceptions of the feasibility of delivering these CBM tools in the school setting and student perceptions of these tools.

Usage Rating Profile-Assessment. Teachers who delivered the CBM tasks to students on their caseloads completed the *Usage Rating Profile-Assessment* (Chafouleas et al., 2012) for each condition. Teachers rated each of the 28-item on a six-point scale (strongly disagree to strongly agree). Each response was then numerically coded: strongly disagree (rated a value of 1), disagree (2), slightly disagree (3), slightly agree (4), agree (5), and strongly agree (6). Results are provided in Table 21.

Table 21

Teacher's Ratings of the Paper-Pencil and E-based Conditions

	N	Mean	*SD*	Range
Paper Pencil				
Acceptability	10	4.20	.41	3.67 – 4.89
Understanding	10	5.07	0.44	4.67 - 6.00
Home School Collaboration	10	2.57	1.29	1.33 - 6.00
Feasibility	10	5.09	.38	4.50 - 5.83
System Climate	10	5.03	0.51	4.25 - 5.75
System Support	10	2.23	0.89	1.00-4.00
E-Based				
Acceptability	10	4.38	.48	3.78 – 5.11
Understanding	10	5.10	0.70	4.00 - 6.00
Home School Collaboration	10	2.37	1.36	1.00 - 6.00
Feasibility	10	5.18	.43	4.67 – 6.00
System Climate	10	5.08	0.64	4.00-6.00
System Support	10	2.53	1.08	1.00-5.00

In the paper-pencil condition, teachers generally agreed that the CBM tasks were acceptable for use (*M* = 4.20), they understood how to administer the CBM tasks (*M* =5.07), administration was feasible in the school setting (*M* = 5.09) and the tasks fit within the context of the school's climate (*M* = 5.03). Teachers varied in their perception of the need for a home-school collaboration when using the CBM tasks, on average teachers reported a home-school collaboration would not be required (*M* = 2.57). Teachers generally reported that additional resources would not be needed to successfully use CBM (*M* = 2.23).

A similar pattern of responding was noted in the e-based condition. The factors of Acceptability (*M* = 4.38), Understanding (*M*= 5.10), Feasibility (*M*=5.18) and School Climate (*M*=5.08) were positively endorsed suggesting that teachers perceived the CBM tasks was understandable, thought their current resources would be sufficient to administer the tool and the CBM tasks fit within the district culture. Although teachers' varied in their reporting, overall the results suggest that teachers' perceived that a Home and School Collaboration (*M* = 2.37) was not required and additional Systems of Support (*M* = 2.53) would not be needed for implementation.

Student Survey Part A: Helpfulness of e-based features. Each participant rated the e-based features as: "helpful," "kind of helpful," or "not helpful" (Table 22). Three students were not administered the video instructions, as they already had a high level of familiarity with Avenue: PM (2011). Their responses were denoted as "NA." Results are displayed in Table 22.

Table 22

Helpfulness of E-Based Features Rated by Students

	Helpful		Kind of Helpful		Not Helpful		NA	
	n	%	*n*	%	*n*	%	*n*	%
Video Directions	21	57%	12	32%	1	3%	3	8%
Practicing	28	76%	7	19%	2	5%	NA	NA
Feedback	26	70%	9	24%	2	5%	NA	NA
Text Movement								
Maze	23	62%	12	32%	2	5%	NA	NA
Slash	26	70%	9	24%	2	5%	NA	NA
Self-Correction	30	81%	6	16%	1	3%	NA	NA
Visual Timer	10	27%	10	27%	17	46%	NA	NA
Motivational Bar	24	65%	9	24%	4	11%	NA	NA

The majority of participants rated most of the e-based features as helpful or kind of helpful: video demonstration (97% of students rated the item helpful or kind of helpful), practicing (95%), feedback (95%), text movement: Maze (95%), text movement: Slash (95%), self-correction (97%), and motivational bar (89%). In contrast to these generally high ratings of helpfulness, the Visual Timer received mixed helpfulness ratings: 27% (*n*=10) of students rated the timer as helpful, 27% (*n*=10), rated the timer as kind of helpful, and 46% (*n*=17) of students rated the visual timer as not helpful.

Student Survey Part B: Condition preference. Each participant (*N*=37) rated if they preferred the delivery of the passages in paper-pencil format, e-based format, or if they had no preference. Similar to Student Survey Part A, three students were not administered the video instructions, as they already had a high level of familiarity with Avenue: PM (2011). Their responses were denoted as "NA." Results are displayed in Table 23.

Table 23

Preferences by Condition (Paper-Pencil verses E-Based) Rated by Students

	Paper-Pencil		Computer		Equal Preference		NA	
	n	%	*n*	%	*n*	%	*n*	%
Introduction								
Directions	3	8%	14	38%	17	46%	3	8%
Practicing	4	11%	17	46%	16	43%	NA	NA
Testing Items								
Maze	6	16%	23	62%	8	22%	NA	NA
Slash	11	30%	16	43%	10	27%	NA	NA
More Fun	5	14%	16	43%	16	43%	NA	NA
Recommend to a friend	3	8%	15	41%	19	51%	NA	NA

The survey included six items: directions, practice, Maze, Slash, enjoyment and recommendation status. For presentation of directions, most participants rated no preference as to the format (46%) or preferred the e-based (38%), a minority preferred paper-pencil (8%). For the practice items, a similar pattern was observed, no preference (43%), prefer e-based (46%), and a minority preferred paper-pencil (11%). For Maze most students preferred the e-based delivery (62%) or had no preference (22%), and a minority preferred paper-pencil (16%). However, for Slash the preference was divided

(30% paper-pencil, 43% e-based, 27% no preference). Participants generally rated the e-based as more fun than paper-pencil (43%) or rated both conditions the same (43%), and a minority thought paper-pencil was more fun (14%). The majority of students would recommend the e-based (41%) or both conditions (51%) to a friend; only 8% recommended the paper-pencil condition.

Fidelity of Administration

Across the nine teachers of deaf and hard of hearing, fidelity of administration across paper-pencil and e-based was 94% (80 to 100%). More specifically, fidelity of administration was 82% to 100% (M = 93%) for paper-pencil and 79% to 100% (M = 95%) for e-based. Due to one student's prior use of Avenue: PM (2011) one teacher did not deliver the demonstration and practice items as scripted but rather provided modified directions, which is reflective of the 79% fidelity score using the e-based. For the teacher who was partially observed, the observed partial session yielded a 100% fidelity score.

My fidelity of administration as rated by the research assistant was 100% for paper-pencil, 94% to 100% (M = 99%) for e-based and 100% for the WJ-III Passage Comprehension subtest (Woodcock, et al., 2001). For the trained graduate assistant, fidelity of administration of the WJ-III Passage Comprehension subtest (Woodcock, et al., 2001) was 100%.

Summary

This study examined whether student scores from CBM tools (Maze and Slash) yielded sufficient reliability and criterion-related validity for students who are deaf and hard of hearing. For Maze, sufficient correlations were present for paper-pencil ("Race"-"Accident" – CMC, CMC-IMC, CMC-IMC/2) and e-based ("Race"-"Around the World"

CMC only). Sufficient correlations were present for “Race,” and “Accident” (select CBM score types) for both criterion measures. The “Pike” passages yielded sufficient correlations with MAP (NWEA, 2003) for select CBM score types but not with the WJ-III Passage Comprehension subtest (Woodcock, et al., 2001). For e-based, sufficient correlations were present for “Race,” “Accident,” and “Around the World” (select CBM score types) for both criterion measures.

For Slash paper-pencil, sufficient correlations only occurred once (“Twelve” – Total Wrong) for the WJ-III Passage Comprehension subtest (Woodcock, et al., 2001) but occurred more frequently for MAP (NWEA, 2003) (“When,” “Twelve,” and “The” passages, select CBM score types). For e-based, sufficient correlations was present for only the “Twelve” passage (select CBM score types) in paper-pencil and e-based.

Additionally, the effects of e-based vs. paper-pencil delivery formats were examined. Differences between paper-pencil and e-based conditions were generally non-significant for Maze; significant differences between conditions for Slash favored the paper-pencil condition.

CHAPTER 5

DISCUSSION

In this study, I explored whether student scores on the CBM measures yielded sufficient reliability and criterion-related validity via two types of media (paper-pencil and e-based) for students who are deaf and hard of hearing. Second, I examined the extent to which the e-based format affected student scores. Below, I describe the research findings within the context of previous research, acknowledge limitations, provide implications to the field of deafness, and discuss directions for future research.

Findings

Maze reliability and criterion-related validity. For alternate form reliability, some correlation coefficients (measuring correct or correct after adjustment for guessing CBM score types) fell in the sufficient range (r >.80, r = .80 & .84; *ps* <.01), with the majority slightly below the sufficient range (r = .61 to .78). The correlations for the score of IMC were generally lower (r's = -.03 to .52), not sufficient and non-significant. Thus, the scores for correct and correct scores adjusted for guessing appear to hold some promise for use with further confirmation needed.

These findings are generally consistent with the previous research that explored the use of paper-pencil CBM Maze tools with students who are deaf and hard of hearing. Chen's (2002) correlation coefficients for CMC and CMC-IMC/2 fell at or slightly below the sufficient range (.77 to .86) and Devenow's (2003) correlations coefficients fell in a similar range (.60 to .82) and were statistically significant (p< .001). Chen (2002) also noted lower correlations for IMC (.30 & .55), which is consistent with findings of this study. Barkmeier and Rose's (2009) correlations varied by CBM form (-.21 to .80), in this study less variation was noted across Maze forms, but rather differences were noted between scores.

For criterion-related validity in the paper-pencil condition, the majority of the correlations were sufficient (CMC, CMC-IMC, CMC-IMC/2, *rs* = .50 to .67; IMC, *r* = -.50) and significant. In the e-based condition, the majority of the correlations were sufficient (CMC, CMC-IMC, CMC-IMC/2, *rs* = .50 to .61; IMC, *r* = -.50) and significant. The measure of incorrect maze choices (IMC) was not sufficient and was inconsistently significant. Thus, this finding suggests that the scores of CMC, CMC-IMC and CMC-IMC/2 appear acceptable to use in both the paper-pencil and e-based conditions and IMC does not appear promising.

In previous studies, student performance on Maze was compared to a comprehensive admissions test (Reynolds, 1985, *r* = .62 & .69), a written language subtest (Chen, 2002, CMC & CMC-IMC/2: *r*'s = .77 to .89), the SAT (Harcourt Brace Educational Measurement, 1995) (Devenow, 2003, CMC & CMC-IMC/2; *r*'s = .72 to .89) and the MAP (NWEA, 2003) (Barkmeier & Rose, 2009, CMC-IMC, *r*'s = -.10 to .91, varied by grade level).

The correlation coefficients obtained in this study were below the correlations obtained by Chen (2002) and Devenow (2003) and near the correlations obtained by Reynolds (1985). The findings from this study are below the correlations obtained by Barkmeier and Rose (2009) at the elementary level (*r* = .80 to .91), near the correlations obtained at the middle school level (*r* =.59 to .91) and fall within the large correlation range at the high school level (*r* = -.10 to .85).

The possible variation between this study and previous studies may be attributed to differences in type of criterion measure used (full battery verses subtest), differences in the specific skills measured by each criterion measure, differences in administration protocols and testing materials, and differences in the sample of students who are deaf and hard of hearing in each study. Even though Barkmeier and Rose (2009) used the MAP (NWEA, 2003), the same tool used in this study, a direct comparison was limited, as correlations were not clustered by grade level for this study. Another factor that possibly influenced correlations coefficients with the criterion is that neither of the criterion measures used in this study were initially designed for use with students who are deaf and hard of hearing, which could have influenced results.

These findings suggest that student performance is generally consistent across Maze reading passages and thus may yield a reliable estimate of student reading skill. Even though some correlations were below what was expected based on previous research, the majority of the correlations fell within the sufficient range, suggesting that the Maze may serve as a general outcome measure and may be an appropriate tool to use for some students who are deaf and hard of hearing.

Slash reliability and criterion-related validity. For alternate form reliability, none of the correlations fell within the sufficient range (r = .39 to .73). For criterion-related validity, the correlations between the WJ-III Passage Comprehension subtest (Woodcock, et al., 2001) and the Slash were rarely sufficient. For the MAP (NWEA 2003), sufficient correlations were noted in both paper-pencil and e-based for the "Twelve" passage, and sufficient correlations were only present for the "When" and "The" passages in the paper-pencil condition. Since the Slash task is a modified reading experience with texts based on set linguistic structures this may explain the lower correlations as the tasks do not match the reading experience of the criterion measures.

Rose et al. (2008) reported alternate form reliability correlation coefficients for Correct Words Identified as .92 and criterion-related validity as .58 to .75. Notable differences between Rose et al. (2008) and this study on the following features are present: sample size (*n* = 101, vs. *n* = 37), demographics, (primarily residential school to primarily mainstream classroom) length and construction of Slash passages (3 min vs. 1 min, isolated vs. graduated passages) and scoring protocol (Correct Words Identified vs. four scoring types). Due to the notable differences, direct comparisons between these two studies are not possible.

Maze: Paper-pencil vs. e-based. In this study, the effect of type of administration was explored. For the Maze, when probes were analyzed individually, the differences between conditions were generally non-significant except for one score (Race passage, IMC score, favoring paper-pencil). When the passages were collapsed, again the difference in mean performance was generally non-significant expect for one score (CMC, favoring e-based). These results suggest that student performance did not vary based on type of administration.

There is no previous research to date comparing the impacts of e-based delivery to student performance on Maze passages with students who are deaf and hard of hearing. However, researchers (Luckner & Bowen, 2010; Luckner, 2013) have suggested that students who are deaf and hard of hearing may benefit from modifications to promote test access. Additionally, researchers (Emary, 2012; Ketterlin-Geller & Johnstone, 2006; Rose & Dolan, 2000) have suggested that e-based delivery might reduce some of the barriers or challenges associated with providing accommodations for students who are deaf and hard of hearing.

The finding that the independent variable (administration type: paper-pencil vs. e-based) did not consistently affect student performance on the Maze could be interpreted in the following ways. The data appear to suggest that if the examiner uses the correct score (CMC) or correct score adjusting for guessing (CMC-IMC, CMC-IMC/2) then either test administration approach (paper-pencil or e-based) appears to hold utility. Adding e-based into the repertoire of CBM delivery options may have potential to increase use of progress monitoring for students who are deaf and hard of hearing.

As multiple modifications were provided in the Avenue: PM (2011) suite, one might hypothesize that students would demonstrate higher performance on the e-based condition as compared to the paper-pencil condition. During the student survey and informal observation of testing sessions, it was noted that students did not consistently use all the modifications provided to them. For example, when viewing the feedback page (students could see the correct and incorrect answer), many students did not appear to actively review their answers prior to moving to the next screen. They were observed to often click “next” within 1 to 2 seconds of seeing the review screen.

Also, during the student survey, some students stated that they had not seen the motivational bar even though it was present on the screen. On the Student Survey, the usefulness of the timer had mixed reviews by students, 27% of students rated the timer as helpful, 27% of students rated kind of helpful, and 46% rated the timer as not helpful. The effects of timer visibility is a current conversation within and outside of the field of deafness (Christ, White, Ardoin, & Eckert, 2013; Derr-Minneci & Shapiro, 1992; Taylor, Meisinger, Floyd, 2013). These informal observations and findings from the Student Survey suggest that there may be a difference between what recommendations are provided and which recommendations are effectively use by students.

According to the Usage Rating Profile – Assessment (Chafouleas et al., 2012), teachers rated both conditions as generally fitting well within the school context. In both conditions, Acceptability (*M* = 4.20 paper-pencil, *M* = 4.38 e-based), Understanding (*M* = 5.07 paper-pencil, *M* = 5.10 e-based), Feasibility (*M* = 5.09 paper-pencil, *M* = 5.18 e-based) and School Climate (*M* = 5.03 paper-pencil, *M* = 5.08 e-based) were positively endorsed. Teachers generally rated that the CBM tasks did not require Home School Collaboration (*M* = 2.57 paper-pencil, *M* = 2.37 e-based) and did not require additional Systems of Support (*M* = 2.23 paper-pencil, *M* = 2.53 e-based). Though not specifically addressed in this questionnaire, it appears that the e-based condition may reduce some barriers of time needed for administration and scoring, which is a concern raised in the previous literature (Luckner & Bowen, 2010; Luckner, 2013).

On the Student Survey students also rated their preference in delivery format. For the Maze, 62% of students preferred the Maze delivered on the e-based only and 16% preferred the paper-pencil condition (22% equal preference). When rating preference with the Maze and Slash combined, 43% of students thought the e-based was more fun than the paper-pencil, only 14% thought the paper-pencil was more fun (43% equal preference). When recommending to a friend 41% would recommend the e-based, only 8% would recommend the paper-pencil (51% no preference). These results suggest that for students with a preference, the e-based was generally preferred.

Overall, these findings suggest that the Maze appears to hold promise for students who are deaf and hard of hearing in either medium. There generally was not a statistically significant difference between the paper-pencil and e-based conditions for the Maze, with a few isolated exceptions. Given some student preference towards the e-based and the possibility that the e-based may save time in administration and scoring, the e-based may be a viable consideration in delivery of the Maze for students who are deaf and hard of hearing.

Slash: Paper-pencil vs. e-based. In this study, the presence of the independent variable (administration type: paper-pencil vs. e-based) was explored for Slash. The reliability correlation coefficients did not fall in the sufficient range and the criterion-related validity data tended to favor the paper-pencil condition. There is no previous research to date comparing the impacts of e-based delivery of testing modifications to student performance on Slash passages with students who are deaf and hard of hearing. Since the testing modifications were generally expected to enhance performance, at first glance the finding of a significant difference favoring the paper-pencil condition seems counter-initiative. Findings can be interpreted in three ways.

First, the Slash task requires students to quickly and accurately place vertical lines between the boundaries between words. Though not directly measured in the study, upon observation of students engaging in the tool, there appeared to be motoric differences between how rapidly a student could locate and draw a vertical line with their pencil as compared to moving the computer mouse to the correct spot and clicking their answer. Some students were observed to have stilted movements when using a mouse or would need to re-click as they had either not placed the computer mouse in the right place on the computer screen or did not click the mouse correctly. These findings highlight that response time may be an important variable influencing a student's score; it is possible that the differences favoring paper-pencil is less of a reflection of the utility of the testing modifications but more a reflection of a student's speed of response.

In the Student Survey, 30% of students preferred the paper-pencil over the e-based and 43% preferred the e-based over the paper-pencil (27% equal preference). This is an interesting finding; as for the Maze 62% preferred the e-based and 16% preferred the paper-pencil (22% no preference). Since students needed to respond less frequently in the Maze (every 7^{th} word) as compared to after each word for the Slash, it appears as if the precision and speed in using the computer mouse may have different relevance in the Slash as compared to the Maze. It is also possible that the frequency or ease of responding may have also impacted student's preference.

Second, passage length appears to be a relevant variable impacting Slash scores. When creating the Slash passage the length of the passage needed to be closely monitored. For Slash, each letter was upper case and there had to be a small space between each letter to allow students space to click between the letters. Additionally, line breaks had to be carefully monitored as the last letter of a line needed to be an end of a word. With these space restrictions, the e-based condition Slash passages needed to be of shorter length. To adequately compare the paper-pencil and e-based conditions, the paper-pencil passages used the same size restrictions when designed.

When administering the tool, many students finished the paper-pencil and e-based Slash text prior to the one min time limit. Even though all scores under one min were prorated, a prorated score cannot be assumed to be exactly the same as the score the student would have obtained if given a longer passage. Even though this limitation likely influenced the paper-pencil and e-based passages the same (as the same size limitations were applied to paper-pencil), there is a concern related to the accuracy of measurement that could have impacted these findings. These findings suggest that limitations on test space may be a relevant consideration that could be a factor that impacts scores.

Third, the findings suggest that when collapsing across scores students who were deaf and hard of hearing identified more correct answers and made less errors when engaging in the paper-pencil condition. As the testing modifications to promote test access were presented as a "package" it is unknown if there were components of the package that enhanced performance and components of the package that deterred performance when students engaged in the Slash. For future research, unpacking the testing modifications may add clarity as to the nuance of what modifications are best for whom, in what combinations, with which CBM measures.

Limitations

Findings of this study must be interpreted in light of the following limitations. First, students varied in demographic characteristics. The age range for the study was large (2^{nd} to 11^{th} grade), the reading level spanned a few grade ranges (2^{nd} to 5^{th} grade), students varied in their hearing loss status (e.g. unilateral, sloping, mild to profound) and varied in their primary language codes and communication modes (e.g. English, American Sign Language, Spanish, Chinese). These inclusive eligibility criteria were used due to the low incidence of deafness; however, these criteria likely created wide variation in the sample and resulted in higher correlation coefficients.

Second, even with extensive recruiting efforts in multiple school districts, the sample size for the study was small (*n* = 37), and even smaller when correlating CBM to the MAP (NWEA, 2003) (*n*= 25). The study meet the minimum sample size to obtain adequate power for some but not all comparisons (minimum of 34 students), the study would have been further enhanced with a larger sample size with learners who were more homogeneous on key characteristics.

Third, as mentioned previously, the study would have been enhanced if students' response time and accuracy using a pencil and computer mouse were quantified. Adding response time would have clarified whether differences between paper-pencil and e-based could be attributed to true differences related to the effectiveness of testing modifications or if another variable (response time) impacted results. Also, having a smaller readability range of the Slash passages and having the opportunity to have more text displayed per passage may have decreased the need for proration.

Fourth, since neither of the criterion-measures were initially designed for use with students who are deaf and hard of hearing, it is unknown if the measures to which the CBM were compared were meaningful indicators of performance. Also, as a student's language proficiency was not directly assessed there is a lack of clarity as to how language proficiency may have affected performance. Even with these limitations, this study provides a starting point for future research in the areas of CBM, deafness and e-based assessment accommodations.

Directions for Future Research

This study built upon previous literature exploring if CBM (Maze and Slash) produce scores with sufficient reliability and criterion-related validity when used with students who are deaf or hard of hearing. It was also the first study to explore the effects of using an e-based medium with deaf and hard of hearing students.

Further research is needed to explore reliability and criterion-related validity with larger sample sizes that have less heterogeneity within the sample (e.g. smaller age range, similar communication modes or reading loss status). These additional research studies could provide insight as to whether there are specific groups of learners to which specific combinations of modifications are most appropriate. Additional research may also resolve inconsistencies in reliability and criterion-related validity coefficients between this study and previous studies.

This study was the first to examine the extent to which an e-based format impacted student scores. In this study, Avenue: PM (2011) was used to deliver a CBM task (Maze) and a CBM-like (Slash) task. All students received the same set (or package) of standardized recommendations. Further research is needed to unpack these recommendations to determine which set of modifications may be most beneficial to students, and/or if different sets of recommendations are needed for students with different sets of demographic characteristics or profiles of need.

Further research is needed to determine the best approach to provide accommodations. Researchers in the field of deafness describe how a barrier to implementation of CBM for students who are deaf and hard of hearing is the time to implement and provide recommendations (Luckner & Bowen, 2010; Luckner, 2013). Research is needed to empirically explore whether both media for administration are appropriate, does the e-based condition yield greater utility in ease of implementation.

Implications to the Field of Deafness

Based on current research in the field, the use of CBM is recommended in the field of deafness to complement data obtained from standardized achievement tests (Rose, 2007). Researchers have also provided recommendations when administering these tools to students who are deaf and hard of hearing (Luckner, 2013). Based on the review of the literature, Maze and Slash rose to the forefront as recommended measures and were used in this study.

Findings from this research study provide additional evidence that CBM may be an appropriate tool for students who are deaf and hard of hearing. For Maze, the results suggest that the scores obtained when students engaged in the Maze were generally consistent with the correlation coefficients obtained for reliability and criterion-related validity in previous studies. The research study also highlighted that providing e-based delivery may be a viable option for Maze, with further evidence needed for Slash.

Conclusion

To increase the literacy outcomes for students who are deaf and hard of hearing, professionals need to provide high quality instruction that is informed by accurate and valid assessment data. This research study explored whether students scores from Maze and Slash had sufficient reliability and criterion-related validity in a paper-pencil and e-based administration format. Overall, findings suggest that Maze may hold promise for use with students who are deaf and hard of hearing in both conditions, with inconclusive results for Slash. Future research is needed to explore the utility of Maze and Slash and the impact of e-based progress monitoring tools.

References

Allen, T. E. (1986). Patterns of academic achievement among hearing impaired students: 1974-1983. In A.N. Shildroth & M.A. Karchmer (Eds.), *Deaf children in America* (pp. 161-206). San Diego, CA: College-Hill Press.

Allinder, R. M., & Eccarius, M. A. (1999). Exploring the technical adequacy of curriculum-based measurement in reading for children who use manually coded English. *Exceptional Children*, *65*(2), 271-83.

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

Appelman, K. I., Callahan, J. O., Mayer, M. H., Luetke, B. S., & Stryker, D. S. (2012). Education, employment, and independent living of young adults who are deaf and hard of hearing. *American Annals of the Deaf*, *157*(3), 264-275.

Avenue: PM (2011). [Computer software]. Minneapolis, MN: Supported by the United States Department of Education: Office of Special Education Programs Stepping Stones Grant.

Barkmeier, L.M. & Rose, S. (2009) *Technical adequacy of the Maze with secondary Deaf and hard of hearing students.* [Unpublished masters thesis.] University of Minnesota, Minneapolis.

Bishara, A. J. & Hittner, J. B. (2015). Reducing bias and error in the correlation coefficient due to nonnormality. *Educational and Psychological Measurement, 75*(5), 785-804. Doi: 10.1177/0013164414557639

Boutin, D. L. (2008). Persistence in postsecondary environments of students with hearing impairments. *Journal of Rehabilitation*, *74*(1), 25-31.

Cawthon, S. W. (2011). Making decisions about assessment practices for students who are deaf or hard of hearing. *Remedial & Special Education*, *32*(1), 4-21. doi:10.1177/0741932509355950

Cawthon, S., & Leppo, R. (2013). Assessment accommodations on tests of academic achievement for students who are deaf or hard of hearing: A qualitative meta-analysis of the research literature. *American Annals of the Deaf,* 158*(3)*, 363-376.

Chafouleas, S. M., Miller, F. G., Briesch, A. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2012). *Usage Rating Profile – Assessment.* Storrs, CT: University of Connecticut.

Chen, Y. (2002). *Assessment of reading and writing samples of deaf and hard of hearing students by curriculum-based measurements* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, MN.

Christ, T. J., & Schanding Jr., G. (2007). Curriculum-Based Measures of Computational Skills: A Comparison of Group Performance in Novel, Reward, and Neutral Conditions. *School Psychology Review*, *36*(1), 147-158.

Christ, T. J., White, M. J., Ardoin, S. P., Eckert, T. L., & VanDerHeyden, A. (2013). Curriculum Based Measurement of Reading: Consistency and Validity Across Best, Fastest, and Question Reading Conditions. *School Psychology Review*, *42*(4), 415-436.

Cohen, J. (1988). Statistical Power Analysis for the Behavioural Sciences. In L. Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

College Board. (1980). *Degrees of Reading Power users manual*. New York: College Entrance Examination Board.

Dennis, A., Bhagwatwar, A., & Minas, R. (2013). Play for Performance: Using Computer Games to Improve Motivation and Test-Taking Performance. *Journal Of Information Systems Education*, *24*(3), 223-231.

Deno, S. (1985). Curriculum-based measurement: The emerging alternate. Exceptional Children, *52,* 219-232.

Derr-Minneci, T. F., & Shapiro, E. S. (1992). Validating Curriculum-Based Measurement in Reading from a Behavioral Perspective. *School Psychology Quarterly*, *7*(1), 2-16.

Devenow, P. (2003). *A study of the CBM MAZE procedure as a measure of reading with Deaf and hard of hearing students* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, MN.

Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying Principles of Universal Design to Test Delivery: The Effect of Computer-Based Read-Aloud on Test Performance of High School Students with Learning Disabilities. *Journal Of Technology, Learning, And Assessment*, *3*(7), 1-32.

Dorn, S. (2010). The political dilemmas of formative assessment. Exceptional Children, 76, 325-337.

Easterbrooks, S. R., & Huston, S. G. (2008). The Signed Reading Fluency of Students Who Are Deaf/Hard of Hearing. *Journal of Deaf Studies and Deaf Education*, *13*(1), 37-54.

Easterbrooks, S., & Stephenson, B. (2006). An examination of twenty literacy, science, and mathematics practices used to educate students who are deaf or hard of hearing. *American Annals of the Deaf*, *151*(4), 385-397.

Edcheckup LLC & Children's Educational Services database (2005) [Computer software and database]. Edina, MN

Emary (2012). Analyzing the various aspects of e-learning modules for the hearing impaired students. *International Journal of Academic Research, 4*(3), 122-130.

Enns, C., & Lafond, L. D. (2007). Reading against all odds: A pilot study of two deaf students with dyslexia. *American Annals of the Deaf, 152,* 63–72.

Espin, C., Deno, S., Maruyama, G., & Cohen, C. (1989, March). *The Basic Academic Skills Samples (BASS): An instrument for the screening and identification of children at risk for failure in regular education classrooms.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review*, *21*(1), 45-59.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children*, *78*(3), 263-279.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Monitoring reading growth using student recalls: Effects of two teacher feedback systems. *Journal of Educational Research, 83,* 103–111.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical

Analysis. *Scientific Studies of Reading*, *5*(3), 239-256.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities*, *45*(3), 195-203. doi:10.1177/0022219412442150

Garson, G. D. (2012). *Significance testing.* Asheboro, NC: Statistical Associates Publishers.

Gilbertson, D., & Ferre, S. (2008). Considerations in the identification, assessment, and intervention process for Deaf and hard of hearing students with reading difficulties. *Psychology In The Schools, 45*(2), 104-120. Doi: 10.1002/pits.20286

Gioia, G. A., Isquith, P. K. & Guy, S. C. (2001). Assessment of executive function in children with neurological impairments. In R. Simeonsson & S. Rosenthal (Eds), *Psychological and developmental assessment* (pp. 317-350). New York: NY: The Guilford Press.

Hammill, D. D., & Larsen, S. C. (1996). *Test of Written Language (3rd ed.)*. Austin, TX: Pro-Ed.

Hammill, D. D., Wiederholt, J. L., & Allen, E. A. (2006). *TOSCRF: Test of Silent Contextual Reading Fluency*, *Examiner's manual*. Austin, TX: PRO-ED.

Harcourt Brace Educational Measurement. (1995). *Stanford Achievement Test, 9th Edition,* Form S (Stanford 9). San Antonio, TX.

Hartmann, E. A. (2010). *Evaluating employment outcomes of adults who are deaf or hard of hearing* (Unpublished Doctoral Dissertation). Teachers College, Columbia University, New York, NY.

Hintermair, M. (2013). Executive Functions and Behavioral Problems in Deaf and Hard-of-Hearing Students at General and Special Schools. *Journal Of Deaf Studies & Deaf Education*, *18*(3), 344-359.

Karchmer, M. A. & Mitchell. R. E. (2003). Demographic and achievement characteristics of Deaf and hard-of-hearing students. In M. Marschark & P. E. Spencer (Eds.) *Oxford handbook of Deaf studies, language and education* (pp. 21-37.). New York: New York: Oxford University Press, Inc.

Kelly, L. P., & Ewoldt, C. (1984). Interpreting nonverbatim cloze responses to evaluate program success and diagnose student needs for reading instruction. *American Annals of the Deaf*, 129(1), 45-51.

Ketterlin-Geller, L. R. (2005). Knowing What All Students Know: Procedures for Developing Universal Design for Assessment. *Journal Of Technology, Learning, And Assessment*, *4*(2),

Ketterlin-Geller, L. R., & Johnstone, C. (2006). Accommodations and Universal Design: Supporting Access to Assessments in Higher Education. *Journal Of Postsecondary Education And Disability*, *19*(2), 163-172.

Kronenberger, W., Pisoni, D., Henning, S., & Colson, B. (2013). Executive Functioning Skills in Long-Term Users of Cochlear Implants: A Case Control Study. *Journal Of Pediatric Psychology*, *38*(8), 902-914.

LaSasso, C. (1980). Validity and reliability of the cloze procedure as a measure of readability for prelingually, profoundly deaf students. *American Annals of the Deaf*, *125*(5)*,* 559-563.

Lederberg, A. R., Schick, B., Webb, M. L., Connor, C. M., Easterbrooks,

S. R., Antia, S., & Kushalnagar, P. (July 2014). The role of language and decoding skills in early literacy development of deaf and hard-of-hearing children. Paper presentation the Society for the Scientific Study of Reading annual meeting, Santa Fe, NM.

Lederberg, A. R., Antia, S., Easterbrooks, S. R., Schick, B., Branum-Martin, L., Connor, C. M., Kushalnagar, P., & Webb, M. L., (April 2014). Development of language and literacy in deaf and hard-of-hearing elementary school children: Strengths and Weaknesses. Paper presentation at the American Educational Research Association annual meeting, Language and Literacy Learning of Deaf and Hard-of-Hearing Children Symposium, Philadelphia, PA.

Luckner, J. L. (2013). Using the Dynamic Indicators of Basic Literacy Skills with students who are Deaf or hard of hearing: perspectives of a panel of experts. *American Annals of the Deaf*, *158*(1), 7-19. doi: 10.1353/aad.2013.0012

Luckner, J., & Bowen, S. (2006). Assessment practices of professionals serving students who are Deaf or hard of hearing: An initial investigation. *American Annals of the Deaf*, *151*(4), 410-417.doi: 10.1353/aad.2006.0046

Luckner, J. L., & Bowen, S. K. (2010). Teachers' use and perceptions of progress monitoring. *American Annals of the Deaf, 155*(4), 397-406.

Luckner, J., Sebald, A., Cooney, J., Young, J., & Muir, S. (2005). An examination of the evidence-based literacy research in deaf education. *American Annals of the Deaf*, *150*(5), 443-456.

Luckner, J. L. & Urbach, J. (2012). Reading fluency and students who are deaf or hard of hearing: Synthesis of the research. *Communication Disorders Quarterly, 33*(4), 230-241.

Luft, P. (2012). Employment and independent living skills of public school high school deaf students: Analyses of the transition competence battery response patterns. *Journal of the American Deafness & Rehabilitation Association*, *45*(3), 292-313.

Luft, P., & Huff, K. (2011). How prepared are transition-age deaf and hard of hearing students for adult living? Results of the transition competence battery. *American Annals of the Deaf*, *155*(5), 569-579. doi:10.1353/aad.2011.0000

McAnally, P. L. & Rose, S. (2012). *Reading Milestones Placement and Monitoring: RMPM Manual.* Austin, TX: PRO-ED.

McGraw-Hill. (1970). *Reading for Concepts.* Columbus, OH: Author.

Marschark, M. (2006). Intellectual functioning of deaf adults and children: Answers and questions. *European Journal of Cognitive Psychology*, *18*(1), 70-89. doi:10.1080/09541440500216028

Marschark, M., & Knoors, H. (2012). Educating Deaf Children: Language, Cognition, and Learning. *Deafness And Education International*, *14*(3), 136-160.

Marschark, M., Spencer, P., Adams, J., & Sapere, P. (2011). Evidence-based practice in educating Deaf and hard-of-hearing children: Teaching to their cognitive strengths and needs. *European Journal of Special Needs Education, 26*(1), 3-16. Doi: 10.1080/08856257.2011.543540

Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What is it and why do it. In M. Shinn (Ed.). *Curriculum-based measurement: Assessing special children.* New York, Guilford Press.

Mayberry, R. I., A., & Lieberman, A. M. (2011). Reading achievement in relation to phonological coding and awareness in deaf readers: a meta-analysis. *Journal of Deaf Studies & Deaf Education*, *16*(2), 164-188. doi:10.1093/deafed/enq049

Minnesota Department of Education (2015). *Students who are deaf and hard of hearing (Fiscal Year 2015): Report to the legislature as required by Minnesota Statues section 125.A.63.*

Mislevy, R. J., Haertel, G., Cheng, B. H., Ructtinger, L., DeBarger, A., Murray, E., & ... Vendlinski, T. (2013). A "Conditional" Sense of Fairness in Assessment. *Educational Research And Evaluation*, *19*(2-3), 121-140.

Moores, D. (2008). Improving academic achievement: Can a response-to-intervention (RTI) model help?. *American Annals of the Deaf, 153*(4), 347-348.

National Center on Student Progress Monitoring (n.d.) *What is progress monitoring?* Retrieved from http://www.studentprogress.org/default.asp

National Reading Panel. (2000). Report of the National Reading Panel. Teaching children to read: An evidenced-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington D.C.: U.S. Department of Health and Human Services.

Northwest Evaluation Association. (2003). *Technical manual for the NWEA Measures of Academic Progress and achievement level tests.* Portland, OR: Northwest Evaluation Association.

Northwest Evaluation Association. (2004). NWEA achievement level test and Measures of Academic Progress. Lake Oswego, Oregon: Northwest Evaluation Association.

Northwest Evaluation Association. (2014). Measures of Academic Progress: An introduction to MAP Computer Adaptive Interim Assessment. Portland, OR: Northwest Evaluation Association.

Oberg, E., & Lukomski, J. (2011). Executive functioning and the impact of a hearing loss: Performance-based measures and the Behavior Rating Inventory of Executive Function (BRIEF). *Child Neuropsychology*, *17*(6), 521-545. doi:10.1080/09297049.2011.555760

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment, 1*(1), 21-21.

Quigley, S.P., McAnally, P., Rose, S., & King, C. (2001) *Reading milestones.* Austin, TX: PRO-ED.

Quigley, S. P., McAnally, P., Rose, S., & Payne, J. (2003). *Reading bridge.* Austin, TX: PRO-ED.

Rhine-Kahlbeck, S. (2004). The assessment of developmental language differences, executive functioning, and social skills in deaf children. Unpublished doctoral dissertation. Gallaudet University, Washington, DC.

Reynolds, H. (1985). Performance of deaf and hard of hearing college students on a criterion-referenced, modified cloze test of reading comprehension. *American Annals of the Deaf. 131*(5), 361-364.

Rose, S. (2007). Monitoring Progress of Students Who Are Deaf or Hard of Hearing. National Center On Student Progress Monitoring.

Rose, D., & Dolan, B. (2000). Universal Design for Learning: Associate Editor's Column. *Journal Of Special Education Technology*, *15*(4), 47-51.

Rose, S., McAnally, P., Barkmeier, L., Virnig, S., & Long, J. (2008). *Silent Reading Fluency Test: Reliability, validity, and sensitivity to growth for students who are deaf and hard of hearing at the elementary, middle school, and high school levels.* (Report No. 9). Minneapolis, MN: University of Minnesota

Russell, M., Hoffmann, T., & Higgins, J. (2009). NimbleTools: A Universally Designed Test Delivery System. *TEACHING Exceptional Children*, *42*(2-), 6-12.

Salend, S. (2009). Using Technology to Create and Administer Accessible Tests. *TEACHING Exceptional Children*, *41*(3), 40-51.

Schirmer, B. R., & McGough, S. M. (2005). Teaching reading to children who are Deaf: Do the conclusions of the National Reading Panel apply?. *Review of Educational Research*, *75*(1), 83-117. doi: 10.3102/00346543075001083

Schirmer, B. R., Therrien, W. J., Schaffer, L., & Schirmer, T. N. (2009). Repeated reading as an instructional intervention with deaf readers: Effect on fluency and reading achievement. *Reading Improvement, 46,* 168–177.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014) *Woodcock Johnson IV Test of Achievement.* Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Woodcock, R. W. (2001). Technical Abstract Woodcock-Johnson III Assessment Service Bulletin No. 2). Itasca, IL: Riverside Publishing.

Shyyan, V., Christensen, L. L., Rogers, C., & Kincaid, A. (2014) *Sign support guidelines for accessible assessments: Insight from cognitive labs.* Minneapolis: MN: University of Minnesota, Enchanced Assessment Grant (#S368A120006), U.S. Department of Education.

Taylor, C. D., Meisinger, E. B., Floyd, R. G., & VanDerHeyden, A. (2013). Variations in Directions and Overt Timing on Oral Reading Accuracy, Fluency, and Prosody. *School Psychology Review*, *42*(4), 437-447.

Thorndike, R. M. & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education: Eighth edition*. New York: Pearson.

Toubanos, E. S. (1995). Review of the Test of Early Reading Ability—Deaf or Hard of Hearing. In Connelly, J. C, & Impara, J. C. (Eds.), *The twelve mental measurement yearbook,* (pp. 1051-1053). Lincoln, NE: Buros Institute of Mental Measurement.

Trybus, R. J., & Karchmer, M. A. (1977). School achievement scores of hearing impaired children: National data on achievement status and growth patterns. *American Annals of the Deaf Directory of Programs and Services, 122,* 62-69.

Wanzek, J., & Vaughn, S. (2010). Tier 3 interventions for students with significant reading problems. *Theory into Practice*, *49*(4), 305-314. doi: 10.1080/00405841.2010.510759

Wayman, M., Wallace, T., Wiley, H., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education*, *41*(2), 85-120.

Woodcock, R. (1987). *Woodcock Reading Mastery Test – Revised*. Circle Pines, MN: AGS.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III.* Itasca, IL: Riverside Publishing.

Appendix A

Maze Passage: Demonstration and Practice

DEMONSTRATION AND PRACTICE

Winter

We have too much rain. Winter is over. Now it is raining. We **(shoe, have, outside)** very cold winters. We have snow **(had, for, them)** five months in the winter. Snow **(out, slowly, is)** fun when winter comes. Snow is **(into, keep, not)** fun after five months.

Appendix B

Maze: "The Accident" Passage

The Accident

Steve started to climb down the ravine toward the fallen tree. It looked **(like, his, too)** an oak tree and Steve was **(sure, can, round)** it would be perfect for the **(campfire, and, two)**. As he climbed down the side **(gave, of, now)** the hill, though, he lost his **(than, first, footing)**, fell forward and twisted his ankle. **(Those, Lying, If)** on the ground in pain, he **(moon, was, big)** not too sure if he could **(like, almost, walk)**. He tried to get up, but **(since, less, the)** ankle would not support his weight. **(Under, Oh, So)** no! Steve thought to himself, how **(away, am, since)** I going to get back? Desperate **(for, since, red)** aid, Steve began calling for help. **(After, However, Has)** several minutes, there was still no **(in, answer, also)**. Steve knew he had to help **(blew, in, himself).** When he spotted a forked tree **(one, to, limb)** about twenty feet away he got **(an, in, are)** idea. He thought he could use **(loud, since, it)** as a crutch!

Note: Children's Educational Services database and Edcheckup LLC (2005)
Used with permission.

Appendix C

Maze: "The Race" passage

The Race

The children had been waiting all morning to go outside and play. After **(but, almost, the)** reading lesson was over, Mrs. Smith **(off, day, led)** the class to the school playground. **(Sat, Either, Everyone)** was having fun playing. But Mike **(slow, dogs, and)** Brad were talking about who was **(the, way, better)** best runner. This gave Mrs. Smith **(as, an, by)** idea. The teacher told the boys **(she, usually, told)** would help them decide who was **(next, now, the)** fastest by setting up a race. **(For, But, Round)** the boys had to agree to **(his, ours, let)** anyone else run, too. Brad and **(both, usually, Mike)** both said yes. Just as the **(since, boys, talking)** were ready to start Kathy said **(for, and, she)** would like to run in the **(feel, first, race)**. All three children started at the **(either, same, hers)** time and ran around the school. **(So, Tells, When)** they were done Brad and Mike **(either, path, knew)** who was the fastest. Kathy was **(now, different, the)** fastest!

Note: Children's Educational Services database and Edcheckup LLC (2005)
Used with permission.

Appendix D

Maze: The "Northern Pike" Passage

Northern Pike

I love to go fishing. Ever since I was a little boy I **(my, have, then)** heard about a great big northern **(off, pike, then)** that lives in the deep part **(later, going, of)** the lake. My dad used to **(tell, me, this)** me about it before I went **(waited, hers, to)** bed. My dad would tell me **(so, about, more)** going out fishing for northern pike. **(He, A, Icy)** would think to himself that maybe **(he, as, from)** would catch the gigantic pike that **(white, out, got)** away from Grandpa many years ago. **(Next, Will, It)** was so big it nearly pulled **(why, but, the)** fishing rod out of his hands. **(It, Is, Next)** pulled so hard that Grandpa could **(from, hers, not)** pull it into the boat. It **(once, before, took)** a big jump and broke the **(would, line, those)**. But Grandpa saw it and said **(only, it, off)** must have been three feet long. **(So, That, Sun)**, every time Dad went fishing he **(very, was, girl)** trying to get the big northern **(pike, different, an)**. So was I!

Note: Children's Educational Services database and Edcheckup LLC (2005)
Used with permission.

Appendix E

Maze: The "Around the World" Passage

Around the World

Have you ever wanted to travel all the way around the world? Most **(since, people, if)** dream of taking such a trip. **(As, If, You**) know what? I actually came close **(to, talking, she)** doing it, and I am only **(gold, twelve, over**) years old. My dad is a **(teacher, always, oh**) and one year he had a **(sooner, beside chance**) to go teach in Hong Kong. **(Well, Our, Sing**) family thought that going to another **(two, across, part)** of the world would be really **(fun, these, able**) and exciting, so we decided to **(neat. inside. go.)** Hong Kong is really a long **(next, way, up**) from the United States. We had **(calmly, to, giving**) fly for almost ten hours to **(hands, yet, get)** there. We lived in Hong Kong **(for, fix, many**) nine months. It was a very **(through, interesting, than**) place. People from all over the **(slow, world, this**) live, work, or travel through Hong **(Or. Kong. Sits.)** After our stay was finished we **(headed, apples, happy)** for home.

Note: Children's Educational Services database and Edcheckup LLC (2005)
Used with permission.

Appendix F

Slash: Demonstration

DEMONSTRATION

T H E H A P P Y B O Y A T E

B I R T H D A Y C A K E

Appendix G

Slash: Practice

PRACTICE

T H E L I T T L E D O G

I S B R O W N

Appendix H

Slash: "When"

W H E N T H E C H I L D R E N W E R E

R E A D Y T H E Y H E L P E D T H E I R

F A T H E R L O O K F O R T H E C A R

K E Y S M O T H E R K I S S E D T H E M

A L L A N D S A I D H A V E A

N I C E D A Y

Note: TOSCRF: 2; Hammill, Wiederholt, & Allen, (2006), Used with Permission.

Appendix I

Slash: "Twelve"

T W E L V E G I R L S W E R E

P L A Y I N G A G A M E A T A

P A R T Y O N T H E W A L L

B E F O R E T H E M H U N G A

P I C T U R E O F A L I O N W I T H

A F I E R C E L O O K I N H I S

E Y E

Note: TOSCRF: 2; Hammill, Wiederholt, & Allen, (2006), Used with Permission.

Appendix J

Slash: "The"

T H E P R I N C E S S H A D A P E T

A N D H E R P E T W A S A T I N Y

G O A T T H E P R I N C E S S H A D

F U N W I T H H E R T I N Y G O A T

Note: RMPM, McAnally & Rose, (2012). Used with permission

Appendix K

Slash: "A"

A G I R L R A N O U T O F T H E

W H I T E H O U S E I N T O T H E

B A C K Y A R D M O T H E R S H E

S A I D M Y L I T T L E P E T B I R D

I S G O N E

Note: TOSCRF: 2; Hammill, Wiederholt, & Allen, (2006), Used with Permission.

Appendix L

E-based Maze: Demonstration

Note: Avenue: PM (2011). Used with Permission

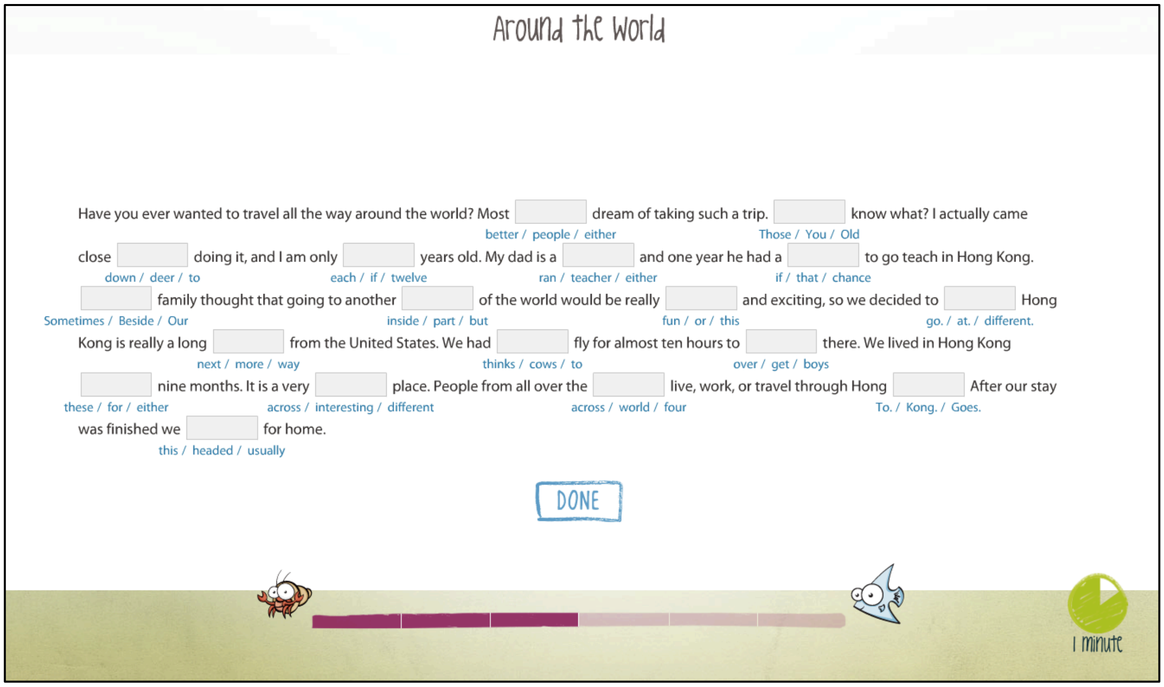
Appendix M

E-Based Maze: Practice

Note: Avenue: PM (2011). Used with Permission

Appendix N

E-based Maze: Testing space

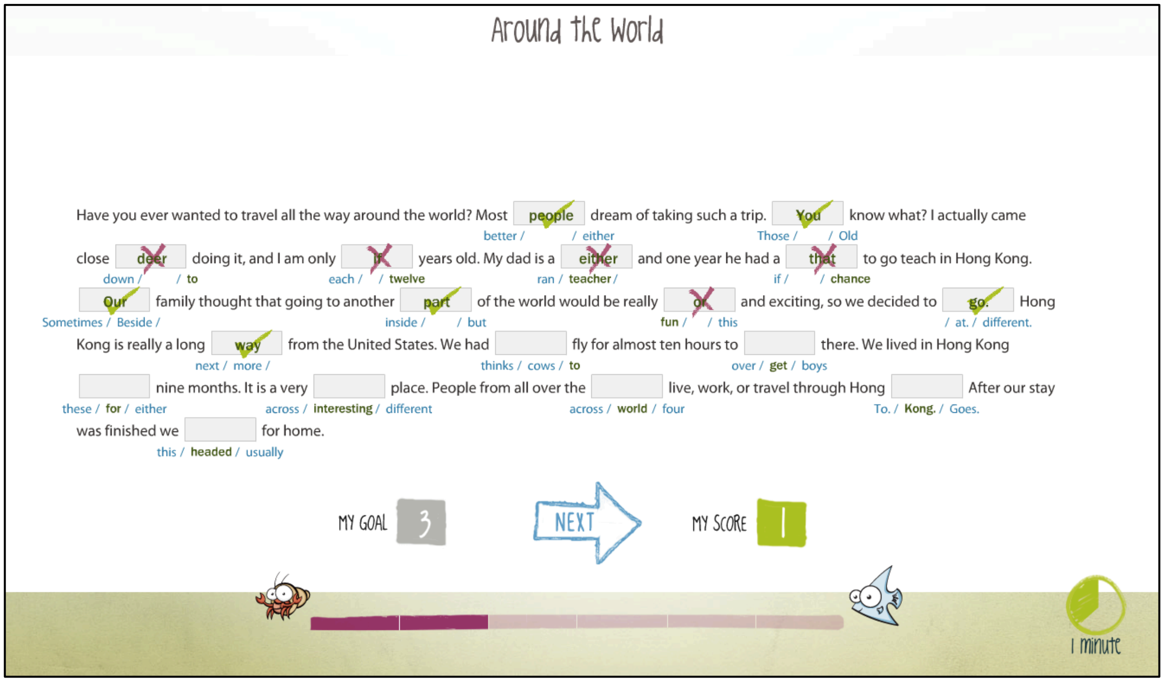


Note: Children's Educational Services database and Edcheckup LLC (2005) and Avenue: PM (2011)

Used with permission.

Appendix O

E-Based Maze: Feedback Page

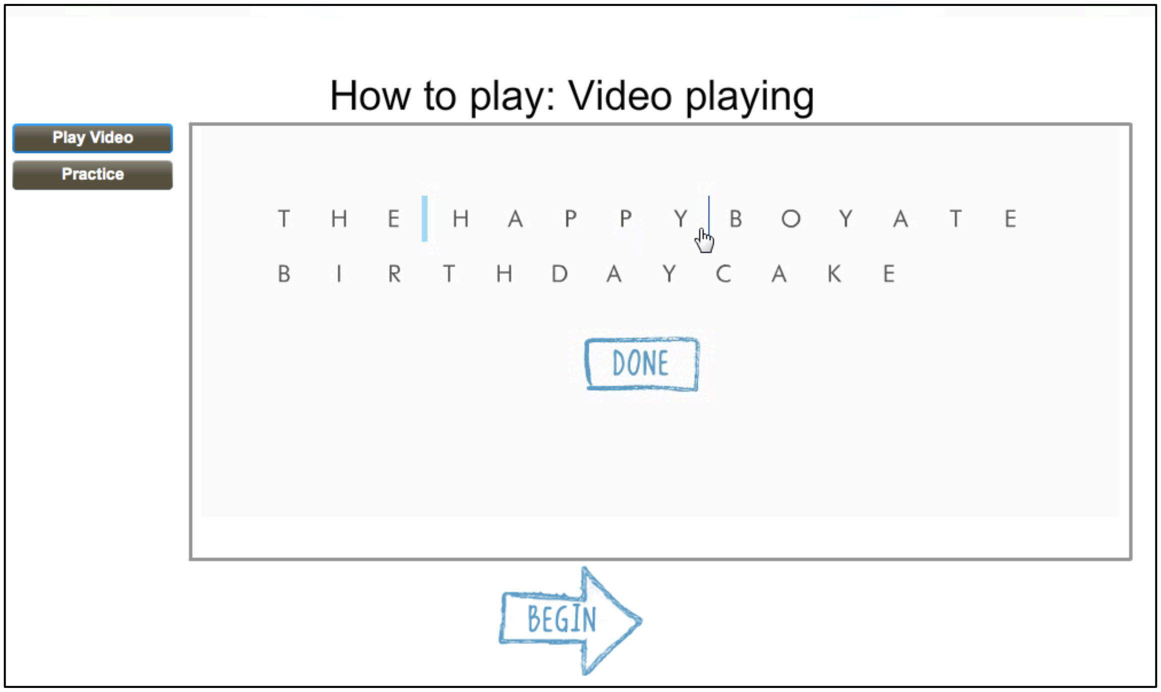


Note: Children's Educational Services database and Edcheckup LLC (2005) and Avenue: PM (2011)

Used with permission.

Appendix P

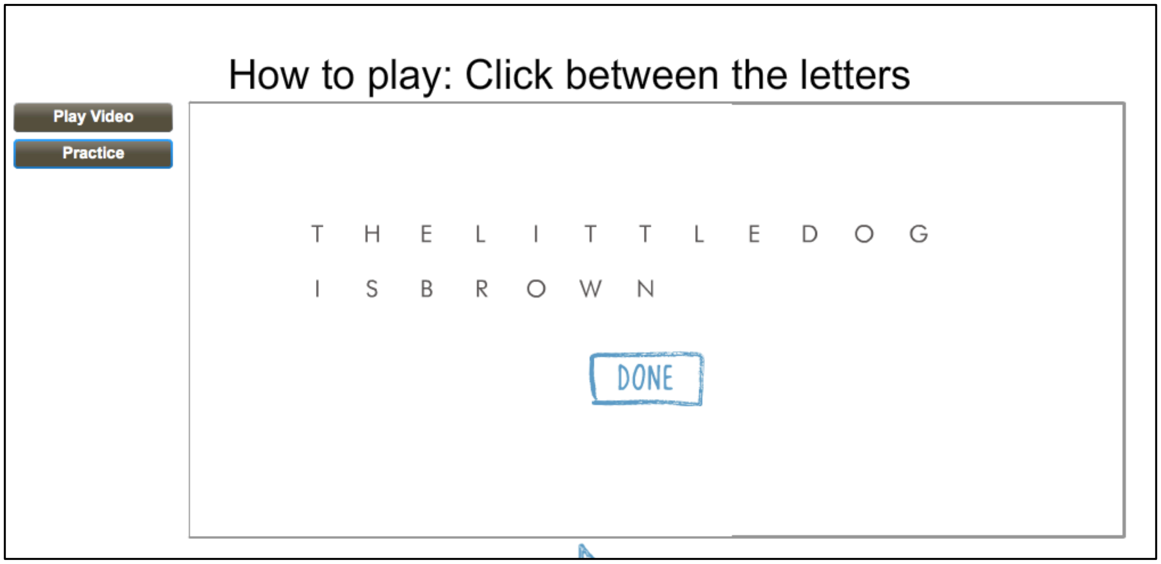
E-Based Slash: Demonstration



Note: Avenue: PM (2011). Used with Permission

Appendix Q

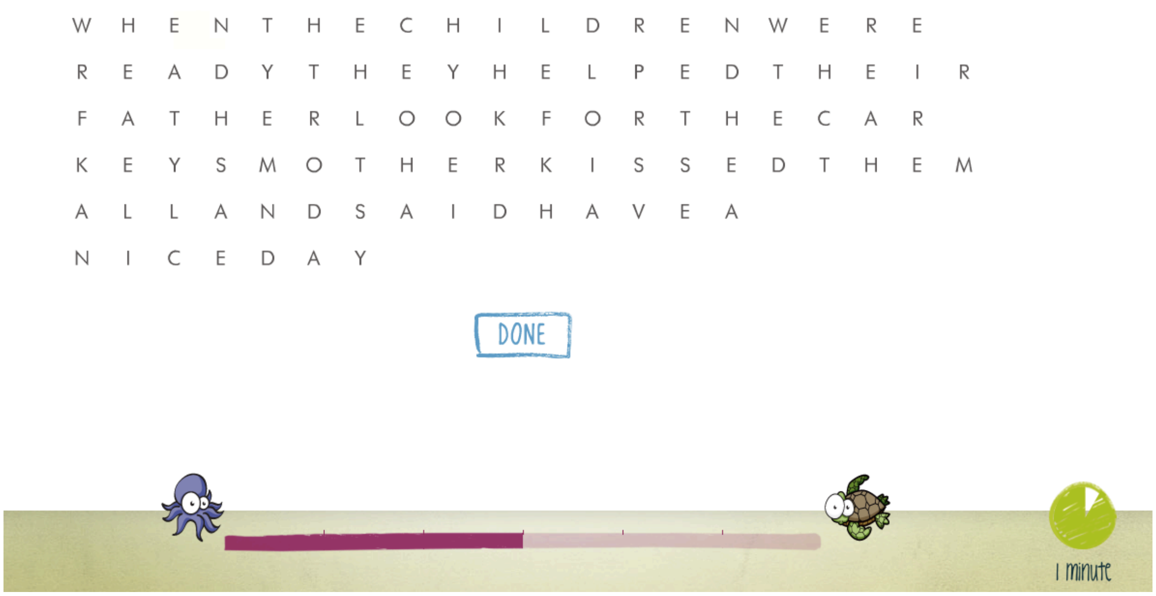
E-Based Slash: Practice



Note: Avenue: PM (2011). Used with Permission

Appendix R

E-Based Slash: Testing space

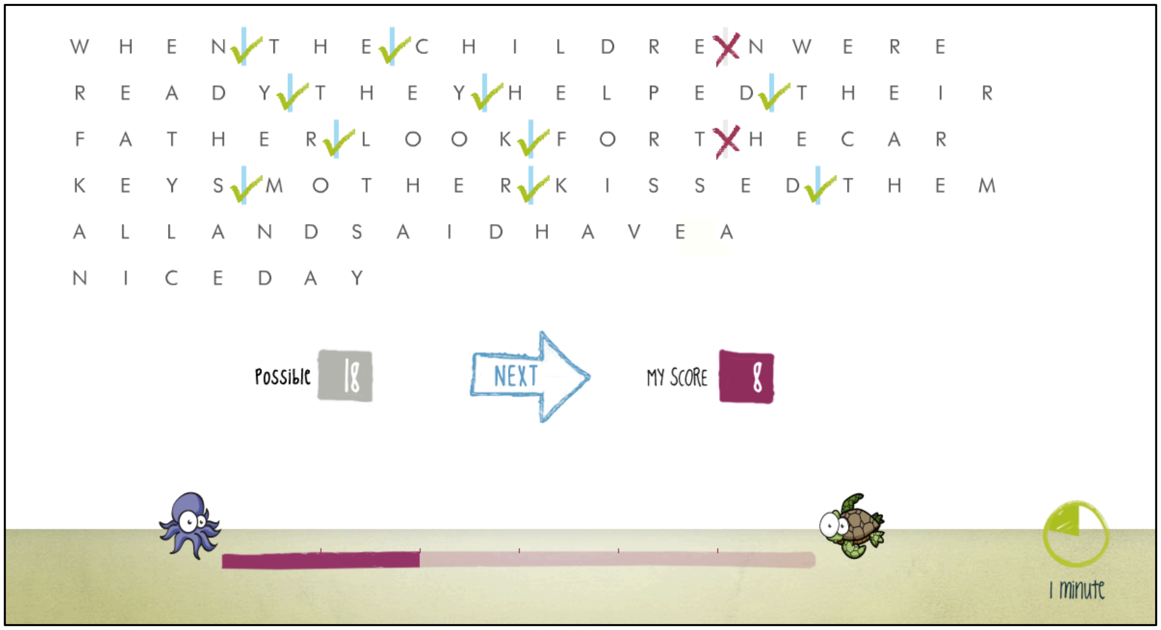


Note: TOSCRF: 2; Hammill, Wiederholt, & Allen, (2006) and Avenue: PM (2011).

Used with Permission.

Appendix S

E-Based Slash: Feedback page



Note: TOSCRF: 2; Hammill, Wiederholt, & Allen, (2006) and Avenue: PM (2011).

Used with Permission.