DOES PASSAGE CONTEXT AND WORD TYPE IMPACT ACCURACY IN MAZE TASKS?

by

STACY-ANN A. BAXTER

(Under the Direction of Scott P. Ardoin)

ABSTRACT

Despite evidence that the maze is a reliable measure of reading comprehension, research suggests great potential for improvement. Specifically, distracters are susceptible to guessing and the syntactic structure of words (i.e., function or content) may affect accurate completion of a maze. The purpose of the current study was to examine variation in student performance as a function of context and the syntactic structure of missing words. Following the administration of an intact maze probe, 225 elementary students were administered, in counterbalanced order, a second intact maze, and a probe with sentences drawn randomly from 3 mazes. Results indicated that students performed significantly better on intact versus randomized sentences, but performance did not differ based upon the syntactic structure of word choices. Findings of this study further question the extent to which mazes measure students' passage comprehension.

INDEX WORDS: maze, reading, comprehension, school-wide assessments

DOES PASSAGE CONTEXT AND WORD TYPE IMPACT ACCURACY IN MAZE TASKS?

by

STACY-ANN A. BAXTER

B.A., University of Richmond, 2006

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2011

DOES PASSAGE CONTEXT AND WORD TYPE IMPACT ACCURACY IN MAZE TASKS?

by

STACY-ANN A. BAXTER

Major Professor:	Scott P. Ardoin
Committee:	Amy L. Reshly Theodore J. Kopcha

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2011

ACKNOWLEDGEMENTS

I would like to extend my sincerest gratitude to the family, faculty, and friends who have supported me throughout this process. To my major professor, Scott Ardoin, thank you for guiding me academically, supporting me, and challenging me intellectually. I would like to thank the students in my program who assisted me in data collection. I would also like to thank Drs. Amy Reschly and T.J. Kopcha for their insight and feedback. Finally, I owe the deepest appreciation to my immediate family for their unfailing emotional, intellectual, and financial support as well as their unconditional love.

TABLE OF CONTENTS

LIST OF TABLES

Page

CHAPTER ONE

INTRODUCTION

Regularly measuring student proficiency and progress in a school's curriculum is an important task that allows teachers to identify the effectiveness of their instruction (Fuchs & Fuchs, 1986). There are multiple ways in which student proficiency and progress can be measured; one of the most common assessment systems used is curriculum-based measurement (CBM). CBM is a system of screening and progress monitoring valid for, among other things, identifying students at risk for reading failure (Deno, 2003). Two of the most common CBM measures are oral reading fluency (CBM-R) and the maze. CBM-R is an individually administered measure of timed oral reading intended to assess global reading achievement. The maze is a group administered task of timed silent reading used to assess students' reading achievement (Fuchs & Fuchs, 1992). Valuable information can be gained by investigating ways in which to improve the maze. To provide readers with a clear rationale for the purpose of this research project conducted to fulfill my master's thesis requirement, the introduction of this thesis provides an overview of curriculum-based assessment (CBA) and CBM, a form of CBA. A review of extant research on the cloze procedure and the maze is then provided. The introduction will conclude with a statement of the purpose of this thesis, along with research-based hypotheses.

Curriculum-Based Assessment

Curriculum-based assessment (CBA) is a set of assessment procedures that measure a student's level of achievement in their program of study (Tucker, 1985). By administering them

multiple times across a school year, schools gain knowledge regarding the progress of their student body, as well as the individual students. Although there are multiple forms of CBA, they share many common characteristics including: (a) measurement methods that target specific skills, (b) administration is short, (c) frequent administration (i.e., 1 – 5 times a week), and (d) resultant data can be graphed (Marston, 1989). Additionally, many CBA measures are used to analyze student performance in the curriculum, select specific target behaviors, and determine the appropriate criteria for skill mastery (Jones, Southern, & Brigham, 1998).

The manner in which a curriculum is sampled to develop CBA materials is one distinguishing feature among the CBA models (Hintze, Christ, & Methe, 2006). There are two identified sampling techniques used for measurement: specific subskill mastery measurement and general outcome measurement (Fuchs & Deno, 1991). Specific subskill mastery measurement models require sampling the curriculum, separating it into a set of subskills, and developing the subskills into a set of hierarchical instructional objectives. Mastery criteria are developed and assessed when appropriate; the focus is on mastery of precise skills. In contrast, the general outcome measurement approach samples the curriculum, but instead of dividing it into subskills, it assesses student proficiency across all skill hierarchies within the curriculum using standardized measures. The focus of general outcome measurement is on global outcomes, instead of mastery of specific skills.

Curriculum-Based Assessment Models

Three CBA models focus on specific subskill mastery measurement: criterion referenced CBA, CBA for instructional design, and curriculum-based evaluation (CBE; Fuchs & Deno, 1991; Hintze et al., 2006; Shinn, Rosenfield, & Knutson, 1989). One primary difference among the various CBA models is the purpose for assessment (Shinn & Bamonto, 1998). Criterion-

referenced CBA uses data collected on classroom skills to make instructional decisions (Blankenship, 1985). The primary purpose is for teachers to gain useful information for planning instruction and choosing instructional resources. Teachers create probes based on a scope and sequence of targeted skills. Then teachers administer probes before instruction begins to identify deficient skills and again after instruction to measure skill mastery. Re-administering criterion referenced CBA probes throughout an academic year allows teacher to assess long-term retention of skills.

CBA for instructional design aligns assessment with instruction, focuses on what students know before instruction, and helps teachers to identify deficits in students' skills (Gickling & Thompson, 1985). This method identifies students' instructional, mastery, and frustrational levels to ensure teachers are designing the curriculum to meet the needs of their students.

The goal of CBE, the third type of CBA used in specific subskill mastery measurement, is to identify the skills students are ready to learn. After an initial criterion-referenced assessment, the specific skills are analyzed to identify deficiencies, enabling assessment to inform instruction. This task analytic model of evaluation allows students to be instructed in and assessed at their present performance level (Howell, 1986).

The fourth type of CBA, CBM, is a general outcome measurement (Fuchs & Deno, 1991). CBM is a set of standardized measurement procedures used to quantify student performance on long-term goals in the academic skill areas of reading, spelling, math computation, and writing (Fuchs, Fuchs, & Hamlett, 1989; Hintze et al., 2006; Hosp & Hosp, 2003). CBM instruments are short and allow for frequent administration, are inexpensive, may be tied to the curriculum, and are sensitive to changes over time (Marston, 1989; Tindal, 1989). CBM is used for a variety of purposes including screening, evaluation, eligibility for special

education, and formative evaluation (Deno, 2003; Marston, 1989). This model was initially developed to provide special education teachers with measurement data to formatively evaluate their instruction and improve their instructional effectiveness (Deno, 2003). Deno (1985) identified CBM as a valid tool for monitoring the progress of students in a curriculum. Instead of simply measuring mastery and moving on, the focus of CBM is on the repeated measurement of global skills (Hosp & Hosp, 2003). Another difference between CBA and CBM is that CBM utilizes long-term assessment procedures that focus on overall achievement in the area being measured instead of assessment and remediation of specific academic skills drawn from the current instructional curriculum (Shinn & Bamonto, 1998). In addition, CBM is employed as an indicator of basic skills (e.g., reading and math) and within a problem-solving approach to help make educational decisions regarding children.

Although CBM exists in several curriculum areas, those used for assessing reading achievement (i.e., CBM-R and maze) are the most widely used and researched group of assessments. Early validation research examining words in isolation, oral reading, and the cloze procedure supported their use in reliably estimating students' reading comprehension proficiency (Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988b). CBM-R, the strongest predictor of reading achievement, measures oral reading rate and accuracy (Faykus & McCurdy, 1998; Fuchs et al., 1988b; Marston, 1989). During CBM-R administrations, students read a passage aloud for 1 min, and the number of words read correctly serves as the dependent measure (Marston, 1989). Extensive research suggests that strong relationships exist between CBM-R and norm-referenced reading achievement tests (Baker et al., 2008; Jenkins & Jewell, 1993; Reschly, Busch, Betts, Deno, & Long, 2009; Ticha, Espin, & Wayman, 2009).

Despite strong evidence of its technical adequacy and its wide use by schools across the country, CBM-R is often criticized for the fact that it requires individual administration (Wesson, King, & Deno, 1984), and lacks face validity as a measure of reading comprehension (Faykus & McCurdy, 1998; Marston, Diment, Allen, & Allen, 1992; Shinn, Good, Knutson, & Tilly, 1992). Even with training on the validity and reliability of CBM-R, research suggests that teachers often doubt its utility as a measure of reading comprehension (Foegen, Espin, Allinder, & Markell, 2001).

The Maze Task

The maze is a timed, modified silent reading and fill in the blank task in which the first sentence of the passage presented is left intact and every seventh word thereafter is deleted and replaced with three word choices (Hosp & Hosp, 2003). One word choice is the target word. Another word choice is a near distracter, which is a word that is the same part of speech as the target word but does not make sense in the sentence. The third word choice is a far distracter that is a different part of speech as the target word and does not make sense in the sentence (Shinn & Shinn, 2002). The maze task can be group administered, as well as administered via the computer, providing more efficient administration than the individually administered CBM-R (Fuchs & Fuchs, 1992).

Validation studies examining the maze suggest its usefulness for a variety of purposes. Jenkins and Jewell (1993) examined the validity of the maze with students in grades 2 through 6 and found moderate correlations (most between .65 and .75) between the maze and two norm-referenced achievement measures. Shin, Deno, and Espin (2000) extended this research by examining the technical adequacy of the maze task for monitoring second grade students' progress and growth. Findings indicated that the maze has strong alternate-form reliability,

reliably assesses student growth across an academic year, and exhibits sensitivity to group and individual student improvement. Comparisons between maze growth rates and CBM-R growth rates suggest that the two measures provide similar growth rates (Fuchs & Fuchs, 1992). Finally, researchers suggest that in fourth grade and above, the maze task may be a better predictor of reading comprehension than CBM-R (Hintze, Shapiro, Conte, & Basile, 1997; Jenkins & Jewell, 1993; Wiley & Deno, 2005; Yovanoff, Duesbery, Alonzo, & Tindal, 2005).

Despite the maze having stronger face validity than CBM-R (Faykus & McCurdy, 1998; Fuchs & Fuchs, 1992), multiple studies exist suggesting that CBM-R explains variance in reading achievement above and beyond the maze (Ardoin et al., 2004; Jenkins & Jewell, 1993). Furthermore, evidence indicates that the maze task does not explain variance in reading achievement above and beyond CBM-R. Given the superiority of CBM-R over the maze, this literature review includes a detailed history of the maze task then describes recognized problems with the maze and the potential resolutions.

History of the Maze

The maze task was derived from the cloze procedure. First introduced by Taylor (1953) as a measure of readability, the cloze procedure was improved upon during the 1970s, renamed the modified cloze, then later modified into the maze task (Parker, Hasbrouck, & Tindal, 1992). A cloze probe is developed from a passage in which every n^{th} word (typically the fifth) is deleted and replaced with a blank line (Helfeldt & Henk, 1985). Students' ability to provide the exact removed word represents their comprehend skills. Later Kingston and Weaver (1970) introduced the first maze task which was in multiple-choice format, with a deletion ratio of 1/5, that included two distracters with one being semantically similar and the other syntactically similar to the deleted word. Researchers argued for the superiority of the maze task over the

cloze procedure, by suggesting that a multiple choice task better mirrored the reading process because it did not require the production of words (Baldauf & Propst, 1979). Empirical research validating the use of the cloze indicated high correlations with reading performance (Baldauf & Propst, 1979; Bormuth, 1968; Smith & Zinc, 1977) and with students' instructional, frustrational, and independent reading levels (Rankin & Culhane, 1969).

Problems with the Cloze and Maze

Despite support for the cloze and maze as measures of reading comprehension, researchers have identified many concerns with these measures. For example, the cloze was criticized for being difficult and frustrating for students and only testing surface level skills (Pikulski & Pikulski, 1977) and for lacking alternate-form reliability (Gillingham & Garner, 1992; Helfeldt, Henk, & Fotos, 1986). In regards to the maze, researchers examined issues with the types of distracters, deletion strategy (e.g., fixed, random, lexical), and relevance of context.

Distracter quality. Parker, Hasbrouck, and Tindal (1992) acknowledged that due to probability, 1/3 of the items on a maze can be answered correctly simply by guessing and half of the remaining items can be answered correctly by focusing on the syntactic structure of words. The validity of the maze task is threatened by the fact that over 60% of a maze can be answered with adequate knowledge of syntax and guessing. Parker et al. offered a solution by organizing distracter types into three categories. Category 1 words are meaningful in the sentence and are the same part of speech as deleted word. Category 2 words are the same part of speech as the deleted word, but are not meaningful in the sentence (i.e., near distracter). Category 3 words are not meaningful in the sentence and are of a different part of speech than the deleted word (i.e., far distracter). Current procedures for developing maze probes require the use of only Category 2 and 3 distracters. Category 1 distracters are the most difficult distracter and the type of

distracter that Parker et al. acknowledge are more frequently absent from maze probes. It seems possible that better distracters, which would require increased attention to the meaning of the passage, will result in the maze predicting greater variance in students' reading comprehension.

Deletion strategy. Research suggests that the fixed deletion ratio used in a cloze task (i.e., every n^{th} word) may result in systematic sampling, potentially resulting in reduced reliability and validity (Helfeldt & Henk, 1985). In an attempt to address concerns about the deletion strategy with cloze measures, researchers explored alternate methods for creating cloze passages such as randomizing the blanks throughout the text and providing a dash for every letter of the target word (Helfeldt et al., 1986). A total random deletion procedure entails removing words at irregular intervals throughout a passage; however, the number of blank spaces between words cannot be controlled. This is problematic because fewer than four words between deletions may not provide sufficient context for a student to accurately complete cloze items (MacGinitie, 1960, 1961 as cited in Helfeldt & Henk, 1985). Attempts to reduce sampling bias by using a random deletion pattern may have adverse effects on the utility of a cloze procedure in effectively discriminating among skilled and poor readers. When Helfeldt and Henk (1985) examined the validity of a randomly deleted cloze versus a traditional cloze, results suggested both measures were comparable, contrary to Taylor (1953), who found a random deletion pattern to be more reliable than a fixed deletion pattern.

Another deletion method examined was the lexical deletion pattern with which the grammatical structure of the target words is the focus of the deletion strategy (e.g., nouns, articles, verbs, and adjectives; Parker et al., 1992). Fries (1952) first conceptualized content words as the words that give sentences meaning and generally include nouns, main verbs, adjectives and adverbs. Words that are not content words are considered function words, which

add little meaning to a sentence. Function word types include articles, prepositions, auxiliary verbs, conjunctions, and pronouns. In support of a lexical deletion strategy, Rankin and Culhane (1969) contended that accurately identifying content words is more difficult because content words depend to a greater extent on the context of a passage. As such, it may be that reducing the number of function words significantly decreases the percentage of correct responses (Tuinman, Blanton, & Gray, 1975).

Kingston and Weaver (1970) hypothesized that deleting every n^{th} word, regardless of word type, would result in the deletion of mostly function words, which measure the structural meaning of language, instead of the context of the passage. To examine this hypothesis, they deleted every fifth content word in a passage and paired each with a distracter. After every five deletions, the 10 words (the five target words and five distracters) were listed in the margins as answer choices. When compared to a traditional cloze and maze, the lexical maze was the single best predictor of scores on a norm-referenced reading achievement test. Despite these findings, existing maze tasks employ the every seventh word deletion pattern without regard to word type.

When examining differences between content and function words with non-native English speakers enrolled in universities, contradictory results were found to those of Kingston and Weaver (1970). For example, Abraham and Chapelle (1992) investigated the difference between function and content words with international students enrolled in an American university. Participants completed one of three types of cloze procedures, two of which were fill-in and one that was multiple-choice. Results indicated that function words were easier, but only within the fill-in cloze task. Kobayashi (2002) examined how function and content words influenced performance of Japanese university students completing the cloze procedure. Students completed a 25 item cloze with every 13^{th} word deleted. Results indicated that function

words were easier than content words. Despite variation in findings across these cloze studies, researchers have thus far failed to examine whether elementary students are more accurate in their choice of function of content words when completing a maze.

The utility of context. Research suggests that comprehension of text may not be influenced by the ordering of sentences (Shanahan, Kamil, & Tobin, 1982). For example, when undergraduates completed a cloze task with the sentences of the cloze placed in random order (randomized cloze) and an intact cloze task, students' performance did not differ. These findings suggest that although understanding the context in which the sentences of a story are placed may assist students in accurately complete a cloze (Shanahan et al., 1982), comprehension may only be required at the sentence level. Other studies, however, suggest that the cloze does require the integration of context across sentences and thus comprehension beyond the sentence level (Bachman, 1982). When comparing the performance of undergraduate students on an intact cloze to a randomized cloze, response accuracy on the intact cloze was superior to those of the randomized cloze (Rye, 1984). A similar effect was observed with a small sample of fifth grade students, whose performance on an intact cloze procedure was significantly greater than on a mixed cloze (McKenna & Layton, 1990). One explanation for the difference in findings across the aforementioned studies was the small sample sizes of the studies. In addition, Rye (1984) and Shanahan et al. (1982) did not expose all participants to both the intact and mixed passage conditions. Despite this question never being resolved completely within the cloze research literature, researchers have yet to examine whether knowing the context of a story contributes to accurately completing maze tasks. The current study addresses these concerns by substantially increasing the sample size and employing a repeated measures design.

Summary

Despite research suggesting that the maze is a reliable and valid measure of reading comprehension, earlier research examining components of the cloze and maze suggests that there is potential to improve the maze (Parker et al., 1992). Specifically, the maze incorporates a fixed deletion ratio (1/7) that includes low quality distracters that are susceptible to guessing. In addition, the syntactic structure of words (i.e., function word or content word) may affect the accurate completion of a maze. Unresolved concerns with the cloze suggest the maze may measure comprehension at the sentence level instead of the paragraph or passage level (Shanahan et al., 1982). Evidence from studies suggests that students can complete cloze tasks without the context of the passage. Finally, with the exception of one study (McKenna & Layton, 1990) involving a small number fifth graders, the utility of context has only been examined with undergraduate students.

Purpose and Hypotheses

In light of the literature evaluating the maze and cloze procedures, the purpose of the current study was to extend the maze literature by examining historic concerns with the cloze that have not adequately been addressed with the maze task. The current study examines variation in student performance as a function of whether sentences were presented in/out of context and the type of word deleted (function verses content words). Specifically, the purpose of this thesis is to extend the literature by conducting research with elementary students rather than undergraduates, providing findings that have greater relevance to the population of students to whom mazes are administered. The current study focuses on the following questions: (1) Do differences exist between students' performance on intact maze and mixed maze probes? and (2) Do differences exist between student performance as a function of the types of word deleted

(function versus content)? It was hypothesized that no performance differences would exist across intact and mixed probes and that students would demonstrate greater accuracy with function words.

CHAPTER TWO

METHOD

Participants and Setting

Participants were 225 students in grades three (n = 81), four (n = 70), and five (n = 74) from one public elementary school in the southeast which services approximately 600 students in grades K-5. Demographic data (i.e., age and gender) were capture for 87% of the participants (via self-report); children ranged in age from 8.66 to 12.08 years (M = 10.30 years, SD = 0.92 years) with 52% of the sample being female. School-wide, approximately 76% of students were White, 12% Hispanic, 4% Black/African-American, and 4% Asian/ Pacific Islander. Approximately 19% of students in the school qualified for free or reduced-priced meals. The experiment occurred in the students' homeroom, in one 20-40 min morning session during instructional time. A single third grade student was excluded from the sample because s/he did not complete all experimental measures.

Measures

All probes were drawn from the fourth grade level AIMSweb maze probes. AIMSweb is a progress monitoring and benchmarking system that uses CBM procedures to assess students' reading and mathematics achievement. AIMSweb provides 33 maze probes at each grade level, the same 33 that are used for CBM-R. The three probes used for fourth grade benchmarking were removed from selection because 4th grade students would have already completed these probes during their fall, winter, and spring benchmarks. In order to identify four probes for use in this study, the percent of deleted words that were function words within each of the remaining

30 probes was determined. Words were classified as a function word if they were prepositions, pronouns, conjunctions, auxiliary verbs, particles, and interjections. After calculating the percentage of function words for each probe, all probes were rank ordered and the five middle probes were selected for inclusion in this study. All five experimental probes ranged from 39.1% to 42.6% function words (M = 40.7%) and contained 24 to 35 sentences (M = 29). See Table 1 for descriptive details regarding each probe. Of the five probes selected, one probe was left intact and utilized as the criterion probe. Each of the remaining four probes (6, 13, 14, and 18) served as both unmodified *intact maze probes* and modified *mixed maze probes.*

Criterion maze probe. The criterion maze probe (5) was chosen from the 30 fourth grade AIMSweb probes previously described. It was not modified in anyway and was presented to students as intended by AIMSweb.

Intact maze probes. When used as intact probes, the probes (6, 13, 14, and 18) were not modified in anyway and were presented to students as intended by AIMSweb.

Mixed maze probes. The four mixed maze probes were developed using the AIMSweb probes that were also employed as intact maze probes (6, 13, 14, and 18). Each mixed maze probe consisted of sentences drawn from three intact maze probes. By drawing sentences from only three of the intact maze probes, participants were administered a mixed maze probe that did not overlap with one of the intact maze probes. For example, students administered intact probe 6 were administered a mixed probe that contained sentences drawn from intact probes 13, 14, and 18. To create a mixed probe, all of the sentences with blanks (deleted words) from the pool of intact probes were numbered, listed, and randomized using a random number string generator (i.e., www.random.org). Sentences were selected and placed in paragraph form to match the corresponding intact probe so that they were visually similar. Additionally, within each mixed

probe, proper names that repeated across sentences were changed to reduce probability that students could gain context from previous or future sentences drawn from the same probe. An example of a mixed probe is provided in Appendix A.

Procedure

One week before data collection, parental passive permission forms (available in Appendices B and C) were sent home to all third, fourth, and fifth grade students in one elementary school. Prior to administration, the five students whose parents opted for them to not participate were provided with an alternate activity such as reading, computer work, or an in-seat activity. Oral assent was gained from the remaining students who were present the day of testing. An example of the script for oral assent is provided in Appendix D. The order of the probe type (i.e., intact probe or mixed probe) within probe sets was counterbalanced to control for order effects.

Students were given an overview of the procedures, provided with directions regarding how to complete the criterion probe, and then allowed 3 min to complete as much of the criterion probe as they could. At the end of 3 min, students were asked to stop, given a follow-up set of directions, and allowed to complete the first experimental probe (intact probe or mixed probe) within their probe set. Students were given 3 min to complete as much of the probe as possible, told to stop at the end of 3 min, switched from pencil to pen, and then received unlimited time to complete the remainder of the probe. After all students finished the second probe, they completed a third and final probe. Procedures for administering the third probe mimicked that of the second probe.

Procedural Integrity

Examiners were school psychology graduate students familiar with standardized testing and maze administration procedures. During one half hour session, examiners were provided with specific training for administering the probe sets described above. A procedural checklist, which included verbatim instructions, was provided during training and for use during data collection, to help ensure procedural integrity. An example of this checklist is provided in Appendix E. Independent observers completed a procedural integrity checklist to evaluate administration fidelity. An example of this checklist is provided in Appendix F. Examiners were aware of the independent observers completing this checklist, but were blind to the results. Study sessions were also audio recorded and used, in conjunction with the procedural checklists, to evaluate procedural integrity. Procedural integrity was evaluated by dividing the number of correctly completed steps by 12, the total number of steps (e.g., time the first probe for 3 min). Procedural integrity averaged 100% across examiners.

Analytic Method

A repeated measures analysis of variance (ANOVA) was conducted to evaluate the effect of mixed and intact probes and function versus content words on performance accuracy for the entire sample and by achievement level (high and low). Participants received the same level probe, thus, one group likely increased statistical power. Each analysis contained two factors; probe type (mixed and intact) and word type (function and content). The dependent measure was the percent accuracy (0 – 100%) of correct words completed within a probe. Main effects (probe type and word type) and interaction effects were examined to test the research hypotheses. A paired samples *t* test was conducted to examine differences in fluency between the intact and mixed probes. The dependent measure was the number of words correct per 3 min (WC3M).

Table 1

Study Measures

Probe	AIMSweb #	% Function Words	# of Words	# Sentences with Blanks
Criterion	4P05	42.55	355	26
Intact 1	4P06	39.58	351	32
Intact 2	4P13	41.67	346	35
Intact 3	4P14	39.13	339	24
Intact 4	4P18	40.43	337	28
Mixed 1	4P13, 14, 18	32.76	406	32
Mixed 2	4P6, 14, 18	36.06	418	35
Mixed 3	4P6, 13, 18	42.72	287	24
Mixed 4	4P6, 13, 14	38.30	303	28

CHAPTER THREE

RESULTS

All variables were examined for normality, kurtosis, and skewness. The Kolmogorov-Smirnova test of normality was significant for the intact and mixed probes, indicating that the data were not normally distributed. Examination of kurtosis and skewness revealed that the intact probes were positively skewed (function words = -3.067; content words = -4.176) with kurtosis that exceeded accepted values (function words = 13.358; content words = 26.834). Efforts were made to normalize the data using statistical transformations, but data remained non-normal. Therefore, data were analyzed with the understanding that not all assumptions were met. Descriptive statistics of student performance on each of the intact and mixed probes are provided in Table 2. A one-way analysis of variance (ANOVA) was conducted for each probe set to examine if differences in student performance were a function of the probe sets to which students were randomly assigned. Results indicated that student performance did not differ based upon the probe sets to which they were assigned (*p*s > .05), suggesting that the probe sets were not significantly different.

A two-way repeated measures ANOVA was conducted to evaluate the effects of probe type (intact and mixed) and word type (function and content) on the accuracy of maze tasks. Wilk's Lambda revealed a significant main effect for probe type, F (1, 224) = 33.394, $p < .001$, with a large effect ($\eta^2_{partial} = .130$), suggesting students completed the intact probes with greater accuracy than the mixed probes. The main effect for word type was not significant F (1,224) = 1.588, $p = .209$, with a small effect ($\eta^2_{partial} = .007$), suggesting no differences between student

accuracy on function words and content words. Finally, the probe type X word type interaction effect was not significant, F (1, 224) = .769, p = .381, with a small effect ($\eta^2_{partial}$ = .003). The lack of a significant interaction effect negated the need to conduct follow-up analyses.

To examine whether reading achievement impacted student accuracy, data from the criterion probe were used to group participants by performance level. Specifically, the number of words correct during the 3 min (WC3M) time limit was used as the dependent measure. All students were divided into 3 achievement groups; high, medium, and low, based on the distribution of scores. For the purposes of this analysis, low performing students scored up to 13 WC3M and high performing students scored 19 or more WC3M. A repeated measures ANOVA was used to examine the effects of probe type (intact and mixed) and word type (function and content) on high and low achieving students. A significant main effect for probe type was observed for low achievers, F (1, 224) = 8.502, p = .005, with a large effect ($\eta^2_{partial}$ = .099), and high achievers, F (1, 224) = 30.198, p < .001, with a large effect ($\eta^2_{partial}$ = .272), suggesting these students completed the intact probes with greater accuracy than the mixed probes. The main effect for word type was not significant for low achievers, F (1,224) = 1.358, p = .247, with a small effect ($\eta^2_{partial}$ = .017), or high achievers, F (1,224) = .247, p = .621, with a small effect ($\eta^2_{partial}$ = .003), suggesting no differences between student accuracy on function words and content words. Finally, the probe type x word type interaction effect was not significant for low achievers, F (1, 224) = 2.125, p = .149, with a small effect ($\eta^2_{partial}$ = .027), or high achievers, F (1, 224) = 1.771, p = .387, with a small effect ($\eta^2_{partial}$ = .021), obviating the need to conduct follow-up analyses.

Finally, a paired-samples t test was conducted to examine differences between students' fluency on the intact and mixed probes. The number of WC3M was used as the dependent

measure. Results indicated that the mean WC3M for the intact probes (M = 23.31, SD = 8.47) was significantly greater than the mean WC3M for the mixed probes (M = 20.13, SD = 6.68), t (223) = 9.50, $p < .001$. Similar analyses examining fluency by word type could not be conducted due to inability to control for the number of content and function words that students were exposed to during the 3 min time limit.

Table 2

Descriptive Statistics by Probe Set

	Intact		Mixed		Function		Content	
Probe Set	M	SD	M	SD	M	SD	M	SD
1 (*n* = 57)	92.62	11.45	89.99	11.54	90.44	13.68	92.05	9.05
2 (*n* = 55)	93.41	10.93	86.71	12.27	90.08	11.93	90.57	10.37
3 (*n* = 56)	92.74	4.70	91.99	7.65	92.40	7.08	92.03	6.02
4 (*n* = 57)	91.90	6.56	91.27	6.88	91.61	5.87	91.89	6.06
Total (*n* = 225)	92.66	8.83	90.01	9.99	91.14	10.14	91.64	8.05

CHAPTER FOUR

DISCUSSION

The maze is widely used by school districts for identifying and monitoring the progress of students struggling in the area of reading. Despite research suggesting the maze is a reliable measure of reading comprehension, research also suggests that there is potential to improve the maze (Parker et al., 1992). Specifically, distracters used as part of the maze are susceptible to guessing and the syntactic structure of some target words (i.e., function words) may make it too easy for students to complete. Unresolved concerns with the cloze, the predecessor to the maze, suggest the maze may only measure comprehension at the sentence level instead of the paragraph or passage level. Finally, with the exception of one study, which employed fifth grade students (McKenna & Layton, 1990), all studies examining the impact of context on individual's ability to complete cloze and maze tasks has involved undergraduates as opposed to elementary students.

The purpose of the current study was to extend the maze literature by examining questions previously addressed within the cloze literature but neither resolved within the cloze literature nor examined within the maze literature. Results of the current study indicated that students completed intact probes with greater accuracy than mixed probes. Differences in students' accuracy with content words and function words were examined, but results revealed no significant differences. Finally, when examining differences in fluency, students demonstrated greater fluency with intact probes than mixed probes.

The current study was the first to examine the impact of probe context on elementary students' performance on a maze task. Current findings replicate and extend the extant literature. Previous research examining the importance of context in regards to students completing cloze tasks is mixed. Some studies suggested that university and fifth grade students' performance on cloze probes was better when having access to content (i.e., intact probe) than without content (i.e., mixed probe; McKenna & Layton, 1990; Rye, 1984) and other studies suggested university students could complete probes at the same level of performance regardless of whether they were provided with the story context (Shanahan et al., 1982). Findings from this study indicated that when elementary students completed intact and mixed probes, they exhibited superior performance on the intact probes. However, despite a significant difference in the percent of correct words between intact (92.66% correct) and mixed probes (90.01% correct), it is questionable whether a difference of only 2.65% is clinically meaningful. The difference is potentially a function of a select few items within the mixed probes requiring comprehension beyond the sentence level, thus decreasing the probability of students answering those items correctly. Nevertheless, the high percentage of correct responses on mixed probes, which in practical terms differ little from the intact probes, suggests that the maze may be only measuring sentence level comprehension as opposed to paragraph or passage comprehension.

The make-up of the mixed probes was designed to prevent students from having any contextual information related to accurately completing items, beyond the information presented within each independent sentence. Thus, in the absence of comprehension beyond the sentence level, students were able to complete the mixed probes with an average accuracy of 90.01%. Specifically, mixed probes contained sentences in paragraph form that were not connected and thus, accurate completion of mixed probes required understanding each sentence individually,

not the passage as a whole. High accuracy on the mixed probes may also be the result of guessing as with only three potential choices, students had a 33% chance of responding correctly. Additionally, previous research acknowledged that by attending to the syntax of words, students can respond accurately to half of the remaining target words.

In addition to accuracy across probes, differences in students' fluency on intact and mixed probes were also examined. Findings indicated that students completed intact probes with greater fluency than mixed probes. Although fluency within maze probes as it relates to this line of research has not been evaluated, prior research suggests that students read words within passages with greater automaticity than word lists because they are able to predict words that come next (Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003). It is likely that students' fluency was interrupted by continual changes in the meaning of the paragraph from one sentence to the next and thus it may be this challenge that resulted in differences in fluency as opposed to differences in the time required by students to choose the correct word choice.

A second purpose of the current study was to examine the impact of word type (i.e., function and content) on students' accuracy on maze probes. Semantically, content words provide meaning to a sentence and function words do not. Specifically, function words explain or create structural and grammatical relationships between content words. Prior research with a multiple-choice cloze procedure suggests that accuracy was greater when choosing function and not content words (Abraham & Chapelle, 1992; Kobayashi, 2002). Current findings contradict those results as the difference between students completion of content (91.64% correct) verses function (91.14% correct) did not differ significantly. Although no differences were found in the current study, it should be noted that the location of the content and function words could not be controlled. That is, not every sentence contained both types of words, nor were content and

function words evenly distributed throughout the probes. In addition, the difficulty of the content words could not be controlled. It is possible that some content words were more difficult than others, leading to a decrease in accuracy.

Limitations

Several limitations should be taken into consideration when interpreting results of this study. First, there was great homogeneity in the sampled population of students. All participants were enrolled in one single, high performing school (i.e., the school has been recognized as falling in the top 10% of all schools in the state, as measured by assessments in reading and mathematics) utilizing the same curriculum. A sample that is more representative of the U.S. population may reveal different results. The homogeneity of the sample likely led to the second and third limitations of the study. Given the high achievement of the students in the study, the large majority of the students were able to complete all items or at least a high percentage of them accurately, resulting in a ceiling effect, particularly for probe sets one and two. The high achievement across students also resulted in the third limitation of the current study; data were not normally distributed. The analyses employed in this study (ANOVA), is typically, however, robust to violations of normality and kurtosis of a distribution may have more of an impact on the F test as skewness. Specifically, large, positive kurtosis values may increase the likelihood of rejecting the null hypothesis; this likely occurred in the current study (Glass, Peckham, & Sanders, 1972). However, it should be noted that significant differences were found between the intact and mixed probes despite small numerical differences. This may suggest that significant differences between intact and mixed probes may also be evident in a normally distributed sample. In addition, even if significant differences were found between function and content words, the extent to which this difference would be meaningful would also be questionable.

A fourth limitation is that all experimental probes were drawn from probes made available by AIMSweb. Other benchmarking systems were not considered for inclusion in this research. However, given the standardization of the maze (e.g., deletion of every seventh word), each benchmarking system creates their maze probes the same way. Therefore, it is likely that maze probes from other publishers will contain similar percentages of function words and content words, and if examined experimentally, will result in similar findings.

Implications

Although the research hypotheses were not supported, results of the current study have several implications for the use of the maze task within schools as a measure of reading comprehension. At most, results of this study suggest that comprehension is at the sentence level for the majority of items presented to students within a maze task. As schools continue to use the maze task as a measure of reading comprehension, these finding should be kept in mind. Schools utilizing the maze likely are also administering CBM-R, which is well-established as a better predictor of reading achievement than the maze (Fuchs, Fuchs, & Maxwell, 1988a; Marston, 1989). Although the maze remains a moderate predictor of reading comprehension, a greater reliance on CBM-R is warranted.

Directions for Future Research

Current findings suggest that improving the maze so that it requires passage level comprehension is warranted. One solution may be to create a maze probe that requires the use of prior context to choose the correct target word. Given the relationship between vocabulary knowledge and reading comprehension, another option would be to create a maze to target specific vocabulary words that could be learned by reading the probe. In this example, accurate selection of the target words would demonstrate comprehension of the passage. Finally, poor

distracter quality was previously acknowledged in the literature (Parker et al., 1992), perhaps a better maze can be developed by changing the type of distracters to decrease the maze task's susceptibility to guessing.

REFERENCES

Abraham, R. G., & Chapelle, C. A. (1992). The meaning of cloze test scores: An item difficulty perspective. *Modern Language Journal, 76*(4), 468-479.

Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L., . . . Williams, K. L. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review, 33*(2), 218-233.

Bachman, L. F. (1982). The trait structure of cloze test scores. *TESOL Quarterly, 16*(1), 61-70.

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., & Beck, C. T. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review, 37*(1), 18-37.

Baldauf, R. B., & Propst, I. K. (1979). Matching and multiple-choice cloze tests. *Journal of Educational Research, 72*(6), 321-326.

Blankenship, C. S. (1985). Using curriculum-based assessment data to make instructional decisions. *Exceptional Children, 52*(3), 233-238.

Bormuth, J. R. (1968). Empirical determination of the instructional reading level. *International Reading Assn Conf Proc Pt 1, 13*, 716-721.

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

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*(3), 184-192. doi: 10.1177/00224669030370030801

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*, 36-45.

Faykus, S. P., & McCurdy, B. L. (1998). Evaluating the sensitivity of the maze as an index of reading proficiency for students who are severely deficient in reading. *Education and Treatment of Children, 21*(1), 1-21.

Foegen, A., Espin, C. A., Allinder, R. M., & Markell, M. A. (2001). Translating research into practice: Preservice teachers' beliefs about curriculum-based measurement. *The Journal of Special Education, 34*(4), 226-236.

Fries, C. G. (1952). *The structure of English*. New York: Harcourt Brace.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*, 488-501.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*(3), 199-208.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*(1), 45-58.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children, 55*(5), 429-438.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988a). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*(2), 20-28.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988b). The validity of informal reading comprehension measures. *RASE: Remedial & Special Education, 9*(2), 20-28.

Gickling, E. E., & Thompson, V. P. (1985). A personal view of curriculum-based assessment. *Exceptional Children, 52*(3), 205-218.

Gillingham, M. G., & Garner, R. (1992). Readers' comprehension of mazes embedded in expository texts. *Journal of Educational Research, 85*(4), 234-241.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research, 42*(3), 237-288.

Helfeldt, J. P., & Henk, W. A. (1985). Usefulness of conventional vs. total random cued cloze tests as measures of reading comprehension. *Journal of Reading, 28*(8), 719-725.

Helfeldt, J. P., Henk, W. A., & Fotos, A. (1986). A test of alternative cloze test formats at the sixth-grade level. *Journal of Educational Research, 79*(4), 216-221.

Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-based assessment. *Psychology in the Schools, 43*(1), 45-56.

Hintze, J. M., Shapiro, E. S., Conte, K. L., & Basile, I. M. (1997). Oral reading fluency and authentic reading material: Criterion validity of the technical features of CBM survey-level assessment. *School Psychology Review, 26*(4), 535-553.

Hosp, M. K., & Hosp, J. L. (2003). Curriculum-based measurement for reading, spelling, and math: How to do it and why. *Preventing School Failure, 48*(1), 10-17.

Howell, K. W. (1986). Direct assessment of academic performance. *School Psychology Review, 15*(3), 324-335.

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*(4), 719-729.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children, 59*(5), 421-432.

Jones, E. D., Southern, W. T., & Brigham, F. J. (1998). Curriculum-based assessment: Testing what is taught. [Article]. *Intervention in School & Clinic, 33*(4), 239.

Kingston, A. J., & Weaver, W. W. (1970). Feasibility of cloze techniques for teaching and evaluating culturally disadvantaged beginning readers. *The Journal of Social Psychology, 82*(2), 205-214.

Kobayashi, M. (2002). Cloze tests revisited: Exploring item characteristics with special attention to scoring methods. *Modern Language Journal, 86*(4), 571-586.

Marston, D., Diment, K., Allen, D., & Allen, L. (1992). Monitoring pupil progress in reading. *Preventing School Failure, 36*(2), 21-25.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it? In M. R. Shinn (Ed.), *Curriculum-Based Measurement: Assessing Special Children* (pp. 18-78). New York: Guilford.

McKenna, M. C., & Layton, K. (1990). Concurrent validity of cloze as a measure of intersentential comprehension. *Journal of Educational Psychology, 82*(2), 372-377.

Parker, R., Hasbrouck, J. E., & Tindal, G. (1992). The maze as a classroom-based reading measure: Construction methods, reliability, and validity. *Journal of Special Education, 26*(2), 195-218.

Pikulski, J. J., & Pikulski, E. C. (1977). Cloze, maze, and teacher judgment. *30*(7), 767-770.

Rankin, E. F., & Culhane, J. W. (1969). Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading, 13*(3), 193-198.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47*(6), 427-469.

Rye, J. (1984). Cloze and intersentential constraint. *Journal of Research in Reading, 7*(2), 113-122. doi: 10.1111/j.1467-9817.1984.tb00261.x

Shanahan, T., Kamil, M. L., & Tobin, A. W. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly, 17*(2), 229-255. doi: 10.2307/747485

Shin, J., Deno, S. L., & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *Journal of Special Education, 34*(3), 164.

Shinn, M. R., & Bamonto, S. (1998). Advanced applications of Curriculum-Based Measurement: 'Big ideas' and avoiding confusion. In M. R. Shinn (Ed.), *Advanced applications of Curriculum-Based Measurement.* (pp. 1-31). New York, NY US: Guilford Press.

Shinn, M. R., Good, R. H., Knutson, N., & Tilly, W. D. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*(3), 459-479.

Shinn, M. R., Rosenfield, S., & Knutson, N. (1989). Curriculum-based assessment: A comparison of models. *School Psychology Review, 18*(3), 299-316.

Shinn, M. R., & Shinn, M. M. (2002). AIMSweb training workbook: Administration and scoring of reading maze for use in general outcome measurement: NCS Pearson, Inc.

Smith, N., & Zinc, A. (1977). A cloze-based investigation of reading comprehension as a composite of subskills. *Journal of Reading Behavior, 9*(4), 395-398.

Taylor, W. L. (1953). 'Cloze procedure': a new tool for measuring readability. *Journalism Quarterly, 30*, 415-433.

Ticha, R., Espin, C. A., & Wayman, M. M. (2009). Reading progress monitoring for secondary-school students: Reliability, validity, and sensitivity to growth of reading-aloud and maze-selection measures. *Learning Disabilities Research & Practice, 24*(3), 132-142.

Tindal, G. (1989). Evaluating the effectiveness of educational programs at the systems level using curriculum-based measurement. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children.* (pp. 202-238). New York, NY US: Guilford Press.

Tucker, J. A. (1985). Curriculum-based assessment: An introduction. *Exceptional Children, 52*(3), 199-204.

Tuinman, J. J., Blanton, W. E., & Gray, G. (1975). The cloze procedure: An analysis of response distributions. *Journal of General Psychology, 92*(2), 177-185.

Wesson, C. L., King, R. P., & Deno, S. L. (1984). Direct and frequent measurement of student performance: If it's good for us, why don't we do it? *Learning Disability Quarterly, 7*(1), 45-48.

Wiley, H. I., & Deno, S. L. (2005). Oral Reading and Maze Measures as Predictors of Success for English Learners on a State Standards Assessment. *Remedial and Special Education, 26*(4), 207-214.

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24*(3), 4-12.

APPENDICES

APPENDIX A: MIXED PROBE 1

Russ (and, was, into) having so much fun that he (didn't, large, dipped) see his little sister, Jenny, sneaking (at, out, up) on him. "Did you have a (huge, from, good) trip?" Aunt Elizabeth asked us. "There they are!" (the, our, Mom) exclaimed. Her Grandma Kate was the best (baker, into, snow) in the entire county. Her mom and (dad, eat, bed) were sitting at the table drinking (kitchen, outside, coffee). At sunset, six hours later, Lucy's (also, men, dad) came to pick her up. We looked out the (forests, bathroom, windows) for our relatives.

James found a (metal, large, nice) spot under an oak tree, far (one, out, in) of view of the house, and (make, sat, she) down. All day long, sugar, flour, butter, (make, could, eggs), and vanilla were all over the (cookies, kitchen, window). My mom and dad, my sister (summer, Rachel, fruit), and I live in Minnesota, and (my, too, us) cousins live in California. Lizzy (tattled, showed, might) Jake how to make big, slow-(always, floated, moving) bubbles. He (filled, water, Sandy) it with warm water and added (so, to, a) generous amount of dish soap. (My, We, To) saw fields and prairies, mountains and (trains, blankets, forests), rivers and valleys, small towns and (big, pass, that) cities.

Mom (took, out, lean) our blankets from her bag. (Tree, Then, First) he lifted it back out, held (on, his, it) at arm's length, and spun around (at, in, it) circles. Kara couldn't (make, they, wait). They would also (give, won, first) everyone on their block a tin (for, of, to) cookies as a Christmas treat. "Are you ready to (bake, their, give) cookies?" Mary asked her grandma as (also, from, she) walked into the house. Grandma Beth (made, after, asked) Sarah a cup of hot chocolate (sugar, before, prize) they began baking.

They would open one (men, for, of) the tins and eat those cookies (after, during, colored) the trip home. We ate meals in (of, the, an) dining car or at the snack (bar, the, side). "Your

seats lean back (to, from, for) sleeping," the conductor told us. Abby knew just what they'd do (also, with, for) the cookies. Next, he dipped (an, at, the) circle into the bucket of water. "Can I go to Grandma's house (past, today, when)?" Lindsey asked as she climbed onto (she, pick, her) mother's lap.

Bill didn't always like (circle, Sandy, please) because she tattled on him whenever (him, her, she) got the chance. Aunt Jane and Uncle David (waved, knew, dozed), and our cousins Jack and Brittany (seats, miles, jumped) up and down. Everyone gave each (other, field, game) a hug. Soon the (travel, rhythm, seats) of the sounds and the rocking (made, help, huge) us sleepy. Grandma Karen was watching for (was, Kendra, men) through the front window of her (could, grand, house) when Ashley's dad dropped her off (it, an, the) hour later. If she (added, told, metal) on him, he might get grounded (for, held, from) a week. When we saw orchards of (sleep, fruit, towns) trees, we knew we were near (my, our, ate) cousins.

APPENDIX B: PARENTAL PERMISSION OPT-OUT LETTER

UGA Department of Educational Psychology and Instructional Technology

Scott P. Ardoin, Ph.D. and Stacy-Ann A. Baxter

Evaluating the Content Validity of Maze Tasks

Parental Permission – Opt-Out

Academic Year 2009-2010

Dear Parents/Guardians,

My name is Dr. Scott Ardoin. Together with one of my graduate students, Stacy-Ann Baxter, we are conducting a research project at your child's school to evaluate the assessments being used in your child school and in districts across the United States. The purpose of the study is to evaluate the quality of predictions of reading comprehension using a Maze task (a reading passage with multiple-choice fill-in-the-blank tasks). A description of the study is provided below.

The project will involve having students complete three maze tasks, while they are in their classrooms. The three tasks will require students to read sentences and circle the word which they believe best fits in the sentence. First, students will be given three minutes to complete a traditional Maze, which will include sentences within paragraphs. Then students will complete a second traditional Maze task, but this one will be untimed. Finally, students will complete another untimed Maze task with sentences that are not placed within a paragraph. The duration of participation should be approximately 30 minutes.

The discomforts involved in this study are minimal, as your child's school administers a maze task to all students on a regular basis. The primary difference between this and other occasions is that we provide the assessment materials and administer the reading passages. In

order to minimize frustration, the study measures will be administered that are on or near the grade level of your child. There are no anticipated risks from participation in this research study. None of the information collected will be used to determine your child's academic grades. To ensure privacy, your child's name will not be associated with his/her responses.

Your child and your child's school may benefit from this study in two ways. First, reading practice can only increase your child's reading abilities. Second, results of this study may assist your child's school and other schools in establishing an accurate, quick measure of reading comprehension.

If you do not wish your child to participate as part of our study, please have your child return the next page to his/her teacher. Participation in this study is voluntary. You may also contact me with any questions that you have about the procedures or purpose of this study by phone (706 542 4110) or e-mail (spardoin@uga.edu). You do not have to allow your child to be in this study if you do not want to. Your child can also refuse to participate or stop taking part at any time without giving any reason, and without penalty or loss of benefits to which she/he is otherwise entitled. If your child does not participate, he/she will have a choice to read silently or complete another academic activity chosen by his/her teacher.

Additional questions or problems regarding your child's rights as a research participant should be addressed to The Chairperson, Institutional Review Board, University of Georgia, 612 Boyd Graduate Studies Research Center, Athens, Georgia 30602-7411; Telephone (706) 542-3199; E-Mail Address IRB@uga.edu.

Sincerely,

_____ _____

Scott P. Ardoin, Ph.D. Stacy-Ann Baxter

APPENDIX C: PARENTAL PERMISSION OPT-OUT FORM

UGA Department of Educational Psychology and Instructional Technology

Scott P. Ardoin, Ph.D.

Evaluating the Content Validity of Maze Tasks

Parental Permission – Opt Out

Academic Year 2009-2010

I have read the above information regarding the study being conducted by Dr. Ardoin and Stacy-Ann Baxter entitled Evaluating the Content Validity of Maze Tasks.

You are not required to return this form. Please only return the form if you <u>DO NOT</u> want your child to participate in the described study. If you do not consent, you may sign and return the bottom portion of this form to the school or e-mail Dr. Ardoin at spardoin@uga.edu.

_ _

Evaluating the Content Validity of the Maze Tasks

Check below if applicable.

____ **I DO NOT** want my child, _____ to participate in this study.

Signature of parent or guardian:_____ Date: _____

APPENDIX D: STUDENT ORAL ASSENT SCRIPT

UGA Department of Educational Psychology and Instructional Technology

Scott P. Ardoin, Ph.D. and Stacy-Ann Baxter

Evaluating the Content Validity of Maze Tasks

Student Oral Assent

To be read by experimenter:

"Hello my name is _____ and I am from the University of Georgia. Today I am doing a research project to see what we can learn from having students read stories and complete sentences. Today you will read 3 stories and circle words that go in blanks. I hope that you will help me. You do not have to do this if you do not want to. How well you do will not affect your class grade. Do you have any questions? Do you wish to participate?"

APPENDIX E: DIRECTIONS AND SCRIPT

1. Start the audio recorder.

2. Read student oral assent: *"Hello my name is _____ and I am from the University of Georgia. Today I am doing a research project to see what we can learn from having students read stories and complete sentences. Today you will read 3 stories and circle words that go in blanks. I hope that you will help me. You do not have to do this if you do not want to. How well you do will not affect your class grade. Do you have any questions? Do you wish to participate?"*

If a student does not assent, inform the teacher so they can complete the alternate activity.

3. Have students take out a pencil while you pass out the probes (face down so they cannot see them) and pens. Instruct student not to begin until you say so.

4. Make sure you have the students' attention (e.g., have the students place their hands in their laps, feet on the floor and eyes on you). Say to the students: "*On your desk are two stories and groups of sentences. Listen to all of my directions before you start. When I say 'Begin' turn your papers over and start reading silently. You will have 3 minutes to read the first story and complete the task. Some of the words in the story are replaced with a group of three words. Your job is to circle the 1 word that makes the most sense in the story. Only 1 word is correct. Work as quickly as you can without making mistakes. Do your best and do not skip any sentences. Keep working until I say 'Stop' or you finish the page. Do not turn the page until I tell you to do so. Do you have any questions?"* (Answer student questions)

5. Say "*Begin*" and start the 3 minute timer.

6. Monitor the students to make sure that they are on task, are circling only one word, are on the correct page, and how fast they are finishing. If a student finishes before the end of 3 minutes, mark the time on the probe.

7. At the end of 3 minutes, say, *"Stop. Put your pencils down and look at me."*

8. After you have everyone's attention, say: *"When I say 'Begin' turn to the next page and start reading silently. At the top of the next page it will either say 'story' or 'mixed sentences'. The story tells about series of events. The mixed sentences are grouped like paragraphs, but they do not go together like a story does. Just like before, while reading you will come to a group of three words, circle the 1 word that makes the most sense. Work as quickly as you can without making mistakes. Do your best and do not skip any. Keep working until I say 'Stop'. When I say stop you will put your pencil down, look at me, and wait for more directions. Do you have any questions?"* (Answer student questions)

9. After saying *"Begin"* start the 3 minute timer.

10. Monitor the students to make sure that they are on task, are circling only one word, are on the correct page, and how fast they are finishing. If a student finishes before the end of 3 minutes, mark the time on the probe.

11. At the end of 3 minutes say: *"Stop. Put your pencils down and look at me"*. Once you have everyone's attention, say: *"Now pick up your pens and complete the rest of the page. Do not go to the next page."*

12. Monitor students to make sure that they are on task, circling only one word and that they do not go on once they finish.

13. After all students have finished, repeat steps 8-12.

APPENDIX F: PROCEDURAL INTEGRITY CHECKLIST

Did the experimenter:

- ☐ 1. Read the student oral assent
- ☐ 2. Read the first set of directions
- ☐ 3. Time the first probe for 3 minutes
- ☐ 4. Monitor students during the first probe
- ☐ 5. Read the second set of directions
- ☐ 6. Time the second probe for 3 minutes
- ☐ 7. Monitor students during the second probe
- ☐ 8. During the second probe, stop all students at three minutes and have them use pens
- ☐ 9. Read the third set of directions
- ☐ 10. Time the third probe for 3 minutes
- ☐ 11. Monitor the students during the third probe
- ☐ 12. During the third probe, stop all students at three minutes and have them use pens