MEASURES OF READING COMPREHENSION: THE EFFECTS OF TEXT

TYPE AND TIME LIMITS ON STUDENTS' PERFORMANCE

Lisa G. Falke, B.A.

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

December 2008

APPROVED:

Jesus Ruiz-Rosales, Major Professor
Shahla Ala'i-Rosales, Committee Member
Michael Fabrizio, Committee Member
Richard Smith, Chair of the Department of
    Behavior Analysis
Thomas L. Evenson, Dean of the College of
    Public Affairs and Community Service
Sandra L. Terrell, Dean of the Robert B.
    Toulouse School of Graduate Studies

Falke, Lisa G. <u>Measures of reading comprehension: The effects of text type and time limits on students' performance</u>. Master of Science (Behavior Analysis), December 2008, 100 pp, 11 figures, references, 48 titles.

Although the importance of reading comprehension is generally recognized, a better understanding of the factors influencing measurement of reading comprehension may impact the ability to assess strengths and deficits. The current study examined the effects of text type and time limits on the rate of students' performance across four common assessments of reading comprehension. Results showed similarities between performance with narrative and expository texts and across time limit conditions for all of the assessments. In terms of comparing across reading comprehension assessments, the findings are limited by the differences in the response channels and stimulus conditions of each assessment. The results have implications for the development of measurement systems and the assessment of reading comprehension.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

The ability to read widens communication and learning opportunities. Text can be used to share information, express desires and further connections with other people. A lack of understanding written text can limit possibilities for social and academic success. In addition, with the recent implementation of the No Child Left Behind Act (NCLB) (2001) schools and teachers are under greater pressure to show high reading comprehension achievement on statewide tests. However, according to statistics presented by the National Center for Education Statistics (NCES) (2007), only 39% of fourth graders across the country read at levels deemed proficient (see U.S. Department of Education, 2007 for a definition) or better and the country's average scale score (on a scale ranging from 0 to 500) has only increased five points since 1992 (U.S. Department of Education, 2007). Since test results have seen little improvement and some students continue to be ineffectual readers, there is still important research to be done with regards to reading comprehension.

There are two components to reading: decoding the text and extracting meaning from what is read. It is important to note that the two components affect each other. It is possible that poor reading comprehension is correlated with poor decoding skills, but research has shown that poor decoding is not the sole source of reading comprehension impairment (Nation & Norbury, 2005; Jenkins & Fuchs, 2003; Storch & Whitehurst, 2002). For example, in Frith and Snowling's (1983) comparison of children with autism, children with dyslexia and typical children, they found that children with autism did not commonly have difficulty with decoding or phonological awareness, but struggled with tests requiring them to use the context of the sentence as a stimulus for

their response, whereas children with dyslexia showed much lower rates of decoding texts than children with autism, but had high scores on tasks where the context of the sentence was important for the response. O'Conner (2004) also differentiated skills of decoding from those of comprehension in his analysis which demonstrated that although children with autism decoded text at the same rates as typical children, they scored lower in retelling the story, identifying the main idea and in skills that require inference. Even though a student may be able to decode the text, there is no guarantee that they will effectively extract meaning. Given the immense importance of understanding written text to an individual's social well being, and the financial contingencies attached to reading performance for school systems, advances in extracting meaning are an ever increasing priority.

Although it is commonly acknowledged that reading comprehension is an important skill for school, job and social success, there is no consensus about what "reading comprehension" means. Sometimes it is loosely defined as "understanding the meaning of a text", but there is no clear operational definition for "understanding the meaning". Reading comprehension is most commonly defined by the many measures used in its assessment (Daly, 2005; Frith, 1983; Kendall, 1980; Lahey, 1973; Myles, 2002; Rasool, 1986). It is possible that a consistent definition is difficult to attain because we are actually discussing a complex set of behaviors, and that each assessment only targets a different piece of a more complex construct. It is often acknowledged in the educational literature that many processes go into comprehension (Jenkins, 2003; Nation & Snowling, 2004; Jenkins & Fuchs, 2003; Lorch et al., 2004; Storch & Whitehurst, 2002; Norbury & Bishop, 2002; Cain, Oakhill & Bryant, 2004;

Ouellette, 2006; Nation et al., 2004), but rarely is it considered that comprehension itself is merely a construct of these many behaviors.

The way reading comprehension is measured will affect the skills that are seen. Both early and recent research has identified that reading comprehension involves multiple components, which are sometimes examined depending on the mode of assessment and the format of the reading materials (Pearson & Hamm, 2005; Storch & Whitehurst, 2002). Therefore, it is necessary to examine the skills involved in each assessment to expose the latent skills being tested. Multiple-choice questions are the most common measure of reading comprehension in research and presented in standardized tests (Jenkins & Fuchs, 2003; Lahey, McNess, & Brown, 1973; Nesi et al., 2006; O'Conner & Klein, 2004; Frith & Snowling, 1983; Norbury & Bishop, 2002), but reading comprehension is also measured through recall (the participant is asked to restate everything they remember after they read a passage), cloze reading activities (the participant fills in blanks found every five to seven words in a reading passage), or maze passages (fill in every fifth or seventh word by choosing one of three choices) (Kendall, 1980; Young, 2005). Another measure of reading comprehension recently introduced to the research is the Sentence Verification Technique (SVT) (Rasool & Royer, 1986). In SVT, the children are presented with four test sentences for each sentence in a passage. These sentence types are original (exact replications of the sentence in the passage), paraphrase (changing words in the sentence without altering meaning), meaning change (changing one or two words in a passage sentence such that the meaning of the sentence is changed), and distracter (consistent with the theme of the passage, but is unrelated to any sentence in the passage). The children are

required to identify whether the sentence they are presented with is information found in the passage or new information. Each of these tests require different skills to respond and often have very different stimuli controlling responding.

Some of the skills most commonly discussed in the literature are working memory, incorporation of background knowledge, attention to context cues, vocabulary knowledge, the knowledge and use of story structure, and the generation of inferences (Lorch et al., 2004; Nation & Norbury, 2005; Jenkins, 2003; O'Conner & Klein, 2004; Nation & Snowling, 2004; Norbury & Bishop, 2002; Cain, 2003; Storch & Whitehurst, 2002; Cain, Oakhill, & Bryant, 2004; Jenkins & Fuchs, 2003; Ouellette, 2006). However, it is clear that all of these skills are not included in all measures of reading comprehension. For example, making inferences and using the context cues present in the assessment may be relevant in the multiple-choice test, but not relevant in the recall test. Performance in the recall test requires skills in working memory and response productions (i.e., vocal speech or writing letters and words). Further, the maze and cloze passages require skills in vocabulary knowledge, working memory, and using context cues from the passage, but do not necessarily require the student to make inferences from the text.

In addition to the skills used to perform the assessment, each assessment often requires a different response. For example, some measures may require the child to write a response where another may ask the child to vocally produce a story. Deficits in the skills needed to respond can also affect the experimenter's ability to infer comprehension. Furthermore, the stimulus conditions across measures for reading comprehension are not uniform. Some of the tests contain written cues for responding,

such as multiple-choice and cloze, and others provide no written or spoken cues for the information, like in the recall test.

Unfortunately, most studies in reading comprehension continue to use only one measure and one response format despite the fact that many studies warn of the limitations in just using one test to assess reading comprehension (Lahey, 1973; Young, 2005; Fletcher, 2006; Pearson & Hamm, 2005; Cutting & Scarborough, 2006, Francis et al., 2006; Nation & Snowling, 1997; Shaire & Leiken, 2004). Fletcher (2006) further reiterates the point by saying that "a one-dimensional attempt to assess reading comprehension is inherently imperfect" (p.324). Without an operational definition of reading comprehension, multiple measures may be necessary to get a full picture of performance and different reading tasks should not be seen as interchangeable measures of a single reading construct. Therefore, a more broad approach to measurement seems justified.

For example, Young (2005) found that performance on recall tests was not correlated to performance on answering questions about the text for a young girl with Asperger's syndrome. During the question test, the participant demonstrated perfect performance in reading comprehension, whereas, the recall test revealed limitations in performance. Although the participant correctly answered questions about main idea and inferential information, during the recall tests she did not make statements of the main idea or mention information inferred from the text, but not explicitly stated.

Similarly, Kendall, Mason & Hunter (1980) compared performance across four types of assessments of reading comprehension (recall, cloze, maze, and multiple-choice questions) with 164 fifth graders. Each student read four passages and was

randomly assigned to a different comprehension condition. They found that students performed significantly worse on the recall test than the other reading comprehension tests and that students performed best on the multiple choice and maze tests. The authors came to the conclusion that none of the measures in isolation provide a measure of reading comprehension, but together, the four assessments provided a more informative picture.

Other variables that affect the comprehension scores across all of the various tests are: The characteristics of the text used for assessment, the ability to reference the text, time limits placed on performance, and the instruction provided prior to the test. One dimension commonly discussed in the literature is the characteristics of the testing materials (Lahey, 1973; Fletcher, 2006; Nation & Norbury, 2005; Cain, 2003; Storch & Whitehurst, 2002). A number of studies have examined the influence of text dimensions on reading comprehension. For example, Young (2005) looked at preference for the text topic as a variable. The author examined recall performance across preferred and non-preferred reading topics and in different reading time requirements. Her results showed that preference for the topic of the reading had little effect on recall performance.

Kendall, Mason & Hunter (1980) used three types of reading materials (expository, narrative, and fairy tale) in their study. The children performed slightly better on fairy-tale passages than narrative passages and slightly worse on expository passages than narrative passages. Even though there was only a slight difference across text types, the performance within text type varied significantly among the four individual passages. The authors came to the conclusion that seemingly comparable passages are difficult to equate and that a larger sample and variety of reading

materials would be necessary to strengthen the comparison. However, the authors'

conclusions are limited by their small text sample size.

Finally, Rasool & Royer (1986) looked at the effect of text type on SVT task

performance. The authors looked at performance across two types of reading texts

(narrative and expository). 44 third graders were presented with four stories, two

narrative and two expository, and a set of test sentences to go with each reading

passage. The authors found that students performed better on the narrative text than

they did on the expository text. However, they mention that the narrative texts used had

a lower readability than the expository texts presented. Also, the results were obtained

with a sample size of only four texts and individual results were not presented for each

passage. In previous studies (Kendall, Mason, & Hunter, 1980; Daly et al., 2005), there

was considerable variability in performance within text types and the averaged together

value presented by Rasool & Royer (1986) may have lost some of that individual

difference which would exaggerate the effect.

Another aspect of assessment that is discussed in the literature is the affect of

applying time limits to performance (Lu & Sireci, 2007; Bridgeman, McBride &

Monaghan, 2004). Timed practice and time-limit testing is used extensively in the

application of standardized tests and in certain formats of instruction, but little research

has been offered that explores the affects of explicit time limits on reading

comprehension performance. Only one study has examined the affect of time limits on

reading comprehension. Lesaux, Pearson, & Siegel (2006) looked at the effects of

extending time limits (providing the students with more time than they needed to

complete the task) on a reading comprehension task for adults with developmental

disabilities. The authors administered the reading comprehension section of the Nelson-Denny Standardized Test to 64 adults with developmental disabilities. They found that all of the participants achieved higher percentile scores when given extra time.

The present study expands on previous research by examining the effects of different assessment tools (recall, cloze, maze, multiple-choice questions and sentence verification), text type (narrative and expository) and time limits on rate of reading comprehension responses. The study adds sentence verification to the range of tests being compared and greatly expands the text samples in the comparison pool. Extending the number of stories presented in each category will help to tease out the individual differences seen within text type. Finally, this study's use of rate to measure reading comprehension behavior will reduce ceiling affects. In this experiment, an alternating treatments design will be used to examine the effects of text type across different tests for reading comprehension. In addition, a reversal will be used to assess the effects of time limits on performance across tests and materials.

METHOD

Participants

Three typically developing children were recruited for this study. All of the participants were enrolled in regular classes and were receiving reading instruction through their respective schools during the time of the study. The participants that were selected had the ability to decode texts up to a fourth grade reading level.

The participants were recruited through a flier (sample in Appendix C) that was distributed to the parents of first through fourth graders at Harvest Christian Academy and posted outside the department of behavior analysis at the University of North Texas. The first three people to contact the principal investigator were selected for the study.

The participants ranged in age from 7 to 10 years old. Two of the participants were siblings, a brother and sister attending a private religious academy. At the time of the study, the male participant, Harry, was 10 years old and attending a fourth-grade general education classroom. His sister, Sally, was 7 years old and attending a first-grade general education classroom. The other participant, Lucy, was an 8-year-old female diagnosed with attention deficit disorder, attending a third-grade public school classroom.

Each participant and at least one of their parents read, signed, and received a copy of the Institutional Review Board (IRB) approved consent form before beginning the study.

Settings

Sessions for all three participants were performed seated at a table. The

participants sat on opposite ends of the same table, while the experimenter stood at an adjacent table with the reading materials. Nothing sat on the participants' table except for timers and writing utensils for each participant. Between tests and between the reading passage and the first test, the participant would walk the finished paper over to the experimenter in exchange for the next paper. In all sessions, the experimenter kept a portable storage container filled with small toys and collectibles under the table where she was standing.

During the read and recall test, the experimenter would sit next to the participant and hold the MP3 player near the mouth of the participant. However, since Harry and Sally had concurrent sessions, the recall portion of the tests was conducted in an adjacent room. One participant and the experimenter went into the adjoining room, out of earshot from the other participant. The experimenter stood across from the participant and held the MP3 player near the participant's mouth.

For Lucy, the sessions were conducted at the kitchen table in either her mother or her father's home. In the kitchen, in addition to the needed materials, there was a computer. The only distractions occurred when a family member would cross through the room to access the kitchen or computer in the next room.

For Harry and Sally, the experimenter administered sessions at the Harvest Christian Academy library or in their home at the kitchen table. In addition to the needed materials, the library contained other tables and bookcases. The only distractions occurred when other students were working at the other tables in the library or when other students were passing through the library to look at books.

Materials

Instructional Passages. The main materials used in this study were 40 narrative

stories and 40 expository stories. The guidelines for passage selection focused on

content, reading perspective and overall structure. A story was categorized as *narrative*

if it was written from the perspective of a character in the story and contained a setting,

an initial event, an internal response, and a consequence or reaction (Rasool & Royer,

2001). A story was categorized as *expository* if it was written from a third-person

perspective or written as descriptive statements and contained general informational

statements and specific examples for support (Rasool & Royer, 2001).

The stories were adapted from the *McCall-Crabb Standard Test Lessons in*

Reading (McCall & Crabb, 1978), *Reading Placement Tests: 3rd Grade* (Scholastic,

2002), *Targeting the TAKS: Reading, Writing, and Mathematics Grade 4* (Steck-

Vaughn, 2005) and *Edhelper.com* (Edhelper, 2007). The stories were all modified to be

between 180 and 315 words and have a readability of 3.2 to 4.5 according to the

Flesch-Kincaid formula. To modify the stories, the experimenter removed sentences

which contained information extraneous to the main idea or plot, rewrote sections to be

more concise, and sometimes added details that did not alter the main idea or plot.

Although the stories were modified, the critical details and main idea of each story were

maintained (see Appendix B for samples of the reading passages).

To ensure that the participants had not contacted the same passages on a prior

occasion, the experimenter showed all materials and material sources to the parents of

the participants and requested that they notify the experimenter if they see anything

their child may have already contacted. The subject matter of the stories spanned social studies, nature, health, animals, science, biographies, and family activities.

In addition to the stories themselves, the experimenter constructed three reading comprehension tests for each story: a cloze passage, multiple-choice questions, and sentence verification (see appendix B for samples of the reading comprehension tests).

Cloze. Removing one word from the story, every 15 to 20 words, and replacing it with a 13-spaced "blank" generated the cloze passage for each story. After removing 10 words from the story and replacing them with "blanks," the experimenter mixed-up the words that had been removed and placed them in a word bank which appeared at the top of the paper (Edhelper.com, 2007; Kendall, Mason & Hunter, 2001; Frith & Snowling, 1983). The word bank displayed the missing words in 2 columns of 5 words each. Each cloze passage contained only 10 blanks and all of them offered a word bank with 10 words from which the participant could select.

Multiple-Choice. All of the multiple choice tests contained 8 questions, each with 4 response choices. The questions themselves were constructed based on the 5 primary reading comprehension objectives set forth in the Texas Assessment of Knowledge and Skills (TAKS) Objectives and Information Packet for 2004: to identify direct information from the text, to predict probable future events or feelings of a character, to identify sequential order of events in the story, to identify the main idea in the text and to identify the meaning of words in the text. The experimenter placed at least one of each type of question on every multiple-choice test. The answers offered for each questions were also based on guidelines stated in the TAKS information packet. The answer choices consisted of one that was consistent with the information in

12

the text, one that was contradicted by information in the text, one that introduces information not offered in the text, and one that is information from the text, but not relevant to the question asked (Texas Education Agency, 1985). The experimenter wrote all answer choices to align to Texas Education Administration (TEA)'s recommendations. The experimenter randomized the order in which the question types and the answer types appeared.

Sentence Verification. For the sentence verification, eight to ten sentences were assembled for each story in accordance with the guidelines set forth in The Sentence Verification Technique (SVT) (Rasool & Royer, 1980). There were four types of test sentences developed: an exact repetition of a sentence from the text (called *an original*), a sentence accurately rephrasing information from the text (called *a paraphrase*), a sentence where one or two words have been changed from the passage sentence, such that the meaning of the sentence has been changed (called *a meaning change*) and, *a distractor,* "a sentence that is similar in syntactic structure to a passage sentence, is consistent with the theme of the passage, but is unrelated to any sentence in the passage" (Rasool, 1980, p. 180). The experimenter included at least one of each type of sentence in every sentence verification test. The order in which the sentence types were presented was randomized, as well as, the amount of each type of sentence was altered for each story.

The experimenter printed each reading passage and each reading comprehension test on its own 8.5" X 11" sheet of white printer paper.

The participants made all responses directly on the 8.5" X 11" sheet of white printer paper containing the reading comprehension test using colored pencils or BiC®

soft-grip pens which were provided by the experimenter.

A small MP3 player was used to record the participants' voices during the recall test. The other materials used included digital timers and small toys and collectibles that were placed in a "prize box."

Measurements

The main measure in this study was rate of responses per minute. For cloze passages, multiple-choice questions, and sentence verification responses were categorized as correct or incorrect. For the read and recall test, responses were categorized as relevant or "other" words recalled.

Cloze Passages. A response was counted as correct when the participant wrote a word in the blank whose meaning completed the statement or sentence while maintaining the same substance as the original passage statement or sentence. If the participant misspelled a word, the word was still counted as correct, as long as it was decipherable and the word carried the correct meaning.

A response was counted as incorrect when the participant left the blank empty or filled in the blank with a word that did not complete the statement in a way that maintained the meaning of the original passage statement.

Multiple-Choice Questions. A response was counted as correct when the participant circled the entire statement or the letter corresponding to the statement of an answer choice that is most accurate and relevant to the information provided in the passage.

A response was counted as incorrect for a question when the participant did not circle an answer, circled more than one answer, or circled an answer that did not

correspond to information provided in the passage or was not of appropriate scope for the story.

Sentence Verification. A response was counted as correct when the participant wrote an "O" next to a statement with information that was found in the passage or an "N" next to a statement containing information that either was not presented in the passage or was different than the information provided by the passage.

A response was counted as incorrect when the participant wrote an "N" next to a statement with information that was found in the passage or an "O" next to a statement containing information that either was not presented in the passage or was different than the information provided by the passage. Also, a response was counted as incorrect when the participant did not write a letter next to the sentence or wrote both an "N" and an "O" next to the same sentence.

Read and Recall. A response was counted as a relevant word when the participant said the word as part of a relevant phrase or a series of relevant phrases. A response was counted as an "other" word when the participant said the word as part of an irrelevant phrase or string of words.

A relevant phrase was a string of words, of any length that portrayed a complete idea and accurately matched the content from the passage the participant just read (See Table 1 for samples). A phrase was considered to be accurate if the general content conveyed information that corresponded correctly, not necessarily exactly, to the reading passage. The exact wording of the participant's phrases did not have to match the exact phrasing in the story in order to be counted as relevant.

Once the relevant phrases were identified, each word in each relevant phrase was counted as a relevant word. Relevant words included the word "and" or other linking words such as "then" or "next" if they were between two relevant complete phrases. Also, in phrases where the general content conveyed information that corresponded correctly to the story, but the name of the character was recalled incorrectly or a word was mispronounced, all the words from the phrase were counted as relevant except the mispronounced word or name. For example, if the participant read, "Martin sat down on the couch," but recalled, "Marvin sat down on the couch," all of the words were counted as relevant words (5 relevant words) except for "Marvin." Further, when the participant used a contraction, the contraction was counted as two words. For example, if the participant said, "It isn't hot inside because there is air conditioning," there were 10 relevant words counted because "isn't" was counted as "is not."

A phrase was considered an irrelevant phrase when it consisted of a string of words that was an incomplete idea or phrase, an inaccurate statement about the passage, a statement about something that did not happen in the passage being tested, or a statement that is completely unrelated to the story (See Table 1 for samples). Phrases that were unrelated to the story included reflective phrases or "thinking" words, such as "um," "I think…" or "And…And…And," as well as, complete phrases regarding information not included in the story like, "I love to read about seals." Also, a phrase that is a direct repetition of a phrase said previously in the recall test or which conveys information that has already been stated during the recall test was considered an irrelevant phrase. Such as, if the participant recalled, "Martin sat down on the

couch...Martin sat down on the couch…" the phrase was counted only once as a relevant phrase and the second repeat was considered an irrelevant phrase. Or if the participant said, "It was hot outside because it was summer. Since it was summer, it was hot," only the first phrase is considered relevant because the second phrase contains redundant information.

Table 1

Sample Sentences for the Read and Recall Test

Sample text: "Martin flopped down on the couch. Then, his dad came into the room."	
Relevant Phrases	Irrelevant Phrases
"Martin flopped down on the couch." 6 relevant words	"The girl sat down on the couch." 0 relevant words, 7 "other" words
"The boy sat down on the couch." 7 relevant words	"The dad sat down on the Couch" 0 relevant words, 7 "other" words
"There was a couch in the room and Martin sat down on it." 13 relevant words	"Martin doesn't like to sit on couches." 0 relevant words, 8 "other" words
"Martin sat down on the…um…the…um… couch." 6 relevant words, 3 "other" words	"Martin sat down on the couch and…and…and someone came into the room" 6 relevant words, 8 "other" words
"Marlin sat down on the couch." 5 relevant words, 1 "other" word	"Let me think…" 0 relevant words, 3 "other" words
"Martin was on the couch and then, his dad came in." 11 relevant words	"Are we almost done?" 0 relevant words, 4 "other" words
"Some boy sat down on a couch." 7 relevant words	"Our couch has a bed inside of it." 0 relevant words, 8 "other" words

Once the irrelevant phrases were identified, each word in each irrelevant string of words was counted as an "other" word. Even if several words from the participant's

recall may have matched the passage, if the general meaning of the phrase was inaccurate with regards to the story, all of the words would be counted as "other" words. Also, if the general meaning of the statement was correct, making the phrase a relevant phrase, but a word was mispronounced or a name was recalled incorrectly, the mispronounced word or name was counted as an "other" word. However, if a word or a name was substituted during a recall test and that word or name changed the meaning of the phrase (i.e. if the participant substituted the name of another character from the story), the entire phrase was considered irrelevant and all of the words in the phrase were counted as "other" words. For example, if the participant recalled "Dad sat down on the couch," no words would be counted as relevant because although a person did sit down on the couch, the participant specified the wrong person in the story and thus, the entire phrase is inaccurate. In addition, linking words ("and," "then," "next," etc.) were counted as "other" words when they did not connect two relevant phrases. For example, if the participant said, "Martin sat down on the couch and then, uh, dad…" only six words (Martin sat down on the couch) were counted as relevant and four words (and, then, uh, dad) were counted as "other" words. Similarly, if the participant had recalled, "Martin sat down on the couch and then his dad left the room," only six words (Martin sat down on the couch) were counted as relevant and seven words (and then his dad left the room) were counted as "other" words because the second phrase is inaccurate. Finally, when the participant recalled "thinking" words in the middle of a relevant phrase, the phrase was still considered relevant and counted as such, but the "thinking" words were counted as "other" words. For example, if the text read, "Martin's father picked up a book and handed it to Martin" and the participant recalled, "Martin's

father picked up a book and…um…and he…and he handed it to Martin," the two phrases were counted as relevant (Martin's father picked up a book and he handed it to Martin), but the repeated "and"s, the "um" and the one repeated "he" are all counted as "other" words.

Rate. In addition to the count of responses, a timer was used to measure the duration of time it took the participant to read the passage and the duration of time it took to complete each test. Once we had a count of responses and the amount of time it took to complete the test, the rate was calculated by dividing the number of responses counted by the number of minutes or fractions of a minute it took to complete the test.

Interobserver Agreement (IOA)

Two graduate students in the Department of Behavior Analysis acted as observers for interobserver reliability. The experimenter held training sessions with each reliability observer separately and all scoring of reliability was done individually, without the experimenter or the other reliability observer able to view the materials or hear the recordings. During the training session, the experimenter provided the reliability observer with instructions and a practice sample for each task. For all of the written tests, the reliability observers were given answer keys and instructed to use a highlighter to mark the answers considered incorrect directly on the participants' original work. For the read and recall test, the experimenter and the reliability observer listened to two sessions together and counted relevant and "other" words, while verbally discussing how the words fit the definition of relevant or "other" words. These two sessions were not calculated as part of the IOA, but were merely used as practice for using the definitions.

Interobserver agreement (IOA) was calculated for 30% of the sessions for each participant. For the cloze passages, the number of times the independent scorer and the experimenter agreed on a correct response was divided by the total number of agreements and disagreements to calculate IOA. The mean IOA for cloze passages corrects responses per minute for 30% of the sessions for each participant was: Harry = 96%, Sally = 99.3%, Lucy = 99%.

For the multiple-choice questions, the number of times the independent scorer and the experimenter agreed on a correct response was divided by the total number of agreements and disagreements to calculate IOA. The mean IOA for multiple choice question correct responses per minute for 30% of the sessions for each participant was: Harry = 100%, Sally = 98.4%, Lucy = 97%.

The number of times the independent scorer and the experimenter agreed on a correct response on the SVT task was divided by the total number of agreements and disagreements to calculate IOA. The mean IOA for corrects responses per minute on the SVT task for 30% of the sessions for each participant was: Harry = 96.4%, Sally = 95.6%, Lucy = 98%.

The principal investigator calculated agreement for relevant and "other" words recalled by dividing the smaller number of recorded instances by the larger number of recorded instances and multiplying by 100 (Poling, Methot, & LeSage, 1995). The mean IOA for relevant words recalled for 30% of the sessions for each participant was: Harry = 88.5%, Sally = 93.6%, Lucy = 86.2%. The mean IOA for "other" words recalled for 30% of the sessions for each participant was: Harry = 84.2%, Sally = 82.6%, Lucy = 86%.

Experimental Design

The design employed in this study was an alternating treatments where, within each session, the condition alternated between consisting of a narrative text and consisting of an expository text for all three participants. The presentation order of the two text types was randomized by the experimenter.

In addition, the tests occurred in either a time limit or a duration testing condition, which ran into a reversal design. In the time-limit condition (A) the participants were given a certain amount of time to finish as much of the test as they could. While in the duration condition (B) the participants were asked to complete the entire test and their work duration was recorded. The sequence of these conditions for the cloze passages, multiple-choice questions, and the sentence verification was BABA. The sequence of conditions for the read and recall test was ABA.

Procedures

All of the sessions took place in the afternoon or evening, after the participants returned from school, after-school activities or day camp. The experimenter ran two sessions per week with each participant. Each session took between 30 and 45 minutes.

Prior to the participants' arrival, the experimenter placed a timer and a writing utensil on the table. During the passage reading and the completion of all reading comprehension tests, the participants sat in a seat at the table, while the experimenter stood and presented materials from an adjacent table. Throughout the session, the experimenter provided general praise statements such as, "You're doing great," "Good work," and "Awesome job," but the comments were not contingent on a response.

During each session, the participant read two passages (one narrative and one expository). At the beginning of each research session the participant was handed one of the two passages. After distributing the reading passage, the experimenter reminded the participant to start the timer when they began to read and stop the timer when they were finished reading. Then, the experimenter said, "Start the timer and read the whole story when I say 'go.' Just try your best. Ready, set, go." The participant would start the timer and begin to silently read the story. When the participant was finished reading the whole story, she stopped the timer and wrote the time at the bottom of the story. Then, she handed the story, with the time written at the bottom, to the experimenter and moved on to the four reading comprehension tests.

For the reading comprehension tests, the participant always completed the recall test first and then, was given the three written tests. The participant received one written test at a time. The order of presentation for the other three written reading comprehension tests was randomized for each session. Once the participant had finished all of the reading comprehension tests for the first reading passage, the second reading passage was distributed. Whether the participant received the narrative passage or the expository passage first was randomly rotated each session. The same procedures were then followed for the second reading passage and its associated reading comprehension tests.

There were two testing conditions for all of the reading comprehension tests: a time-limit test condition and a duration condition.

Time Limit Test Condition. During the time-limit-test condition, there was a time limit presented for each test. Rather than completing the entire test, the participants

were encouraged to recall as much information as they could or do as many items as they could on the test before the timer sounded.

During the recall test, the experimenter sat next to the participant at the table, cleared the timer and opened a new file on the MP3 player. Then, the experimenter set the timer and pushed the record button on the MP3 player. The participants were given 20 seconds to recall as much information from the passage as they could. The experimenter said, "I'm going to set the timer for 20 seconds. Now tell me everything you remember from what you just read until the timer goes off again. Just try your best." Then, the experimenter set the timer for 20 seconds, pushed the record button on the MP3 player and said, "Okay, tell me what you read about." On the first words recalled by the participant, the experimenter pushed the start button on the timer. During the recall, the experimenter held the MP3 player close to the participant's mouth and looked down at the timer. The participant talked directly into the MP3 player. When the timer began to beep, the experimenter held it close to the MP3 player and asked the participant to stop. When the participant was finished, the experimenter hit the record button again on the MP3 player. The recall was saved on the MP3 player for later review. At the end of the recall, the experimenter returned to an adjacent table while the participant remained at the table to complete the written reading comprehension tests.

During the written reading comprehension tests, the participant sat at the table and completed each worksheet while the experimenter stood at an adjacent table and provided a new test each time one was completed. First, the experimenter handed the participant one of the written tests. Directions for completing the test were written across the top of the page (for samples see Appendix B). Prior to beginning each test, the

participant was encouraged to read the directions and ask the experimenter any question they might have in regards to testing procedure. During the time-limit condition, the experimenter had also written a time limit at the bottom of each written reading comprehension test. The experimenter reminded the participant to look at the bottom of the page for their time limit and set the timer for the specified time. Each participant set a timer for the time specified at the bottom of the test. The experimenter said, "When I say 'go,' I want you to try and finish as much of the worksheet as you can before the timer goes off. Just try your best and skip the ones that are too hard." When the participant picked up the writing utensil and indicated that they were ready, the experimenter singled to the participant to start the timer and begin the test by saying, "ready, set, go." The participant started the timer and then, recorded all responses directly on the worksheet. When the timer sounded, the participants were instructed to draw a line under the last question, blank, or sentence they had attempted and hand the paper to the experimenter. When the participant was finished, the test was handed to the experimenter and the experimenter gave the participant the next test. The same procedures were repeated for all three written reading comprehension tests.

Duration Condition. For the duration condition, procedures were the same as the time-limit testing condition except that there was no time limit on any of the reading comprehension tests. The participants were encouraged to tell the experimenter as much as they could remember from the story or answer as many items as they could on the written tests.

During the recall test, in the duration condition, the experimenter said, "I want you to tell me everything you remember from the story and just let me know when you're

finished by saying 'That's it' or 'The End'. Just try your best." Then the experimenter

said, "Okay, tell me what you read about." The participant spoke directly into the MP3

player and when the participant said, "That's it," "The end," or "That's all," the

experimenter stopped the timer, said the time into the MP3 player and pressed the

record button on the MP3 player to stop the recording.

For the written reading comprehension tests, during the duration condition, there

was no time limit written at the bottom of the tests. Instead, the participants were

instructed to finish as many items on the test as they could. The experimenter said,

"Just do your best and try to finish the whole worksheet. Skip any ones that are too hard

and write how long it takes you at the bottom of the paper." Prior to the student

beginning the worksheet, the experimenter reminded the participant to clear the timer,

start the timer before beginning the worksheet and stop the timer when finished. Then,

the experimenter said, "You're going to hit start on the timer and begin when I say 'go.'

Ready, set, go." After the participant had attempted all of the questions, blanks, or

sentences, the timer was stopped, the time was written at the bottom of the page and

the paper was handed to the experimenter.

At the end of each session, the participant was allowed to select a "prize" from

the prize box. The experimenter said, "Thanks for working so hard today. Why don't you

go and pick something from the prize box."

First 20 Seconds Analysis. Using the MP3 recordings of the recall sessions from

the duration condition, the experimenter counted the number of relevant and "other"

words recalled during the entire duration test for each duration session. Then, the same

recording was used to measure the number of relevant and "other" words recalled in the first 20 seconds of those same duration tests.

Recording Procedures

Following each session, the experimenter scored the reading comprehension tests and recorded the scores directly on a standard celeration chart that had been created for each participant, for each test.

During the session, all of the participants' responses were recorded directly on the written reading comprehension tests or, in the case of the recall test, recorded directly onto an MP3 file. Also, the duration to complete each test was written on the bottom of the test page. For the read and recall tests, the experimenter would verbally note the amount of time into the MP3 player at the end of each recall test.

For scoring the cloze passages, multiple-choice questions, and the sentence verification test, the experimenter used a printed answer key to compare the participant's responses to the correct responses for each test. For each test, the experimenter compared the participant's responses with the responses on the answer key and counted the number of correct and the number of incorrect responses. The number of correct and incorrect responses were immediately converted to rate and graphed on the corresponding standard celeration chart.

For the recall test, immediately following the session, the experimenter would listen to the voice recording, once to count the relevant words recalled and then, a second time to count the "other" words recalled by the participant. The experimenter had a dated data sheet where she would immediately record the number of relevant and "other" words recalled for each text and the amount of time the recall lasted. Then, the

number of relevant and "other" words were converted to a rate measure and graphed directly on the standard celeration chart.

All of the scores were graphed as rate of response. Thus, prior to graphing the score, the information had to be converted into a rate measure. First, the durations to complete each test were converted to minutes. Then, to obtain rate, the experimenter divided the number of responses for each test by the amount of minutes or fractions of a minute it took to complete each test. The experimenter would immediately record the participant's scores for each test directly on the standard celeration chart.

RESULTS

Reading Test Comparison

Figures 1 and 2 show Harry, Sally, and Lucy's frequency distributions of correct and incorrect responses for the cloze passages (far left), multiple-choice questions (middle left), the sentence verification (middle right), and read and recall tests (far right) across narrative and expository texts. For each graph, the correct responses or relevant words per minute are presented on the left-hand side of the graph, while the incorrect responses or "other" words per minute are shown on the right-hand side of the graph. The responses are displayed as a count per minute and show horizontally, how frequently a specific score occurred in each condition. The "N"s on the graph represent the number of sessions that occurred in each condition.

For Harry (top row, Figure 1), regardless of the text condition, the highest rates of both correct and incorrect responses per minute occurred in the read and recall test (Harry's median correct response=165, Harry's median incorrect response=12); whereas the lowest rates of correct and incorrect responses per minute were found in the cloze passages (Harry's median correct response=9, Harry's median incorrect response=0).

For Sally (bottom row, Figure1), regardless of text condition, the graphs show that the highest rates of both correct and incorrect responses per minute occurred in the read and recall test (Sally's median correct response= 98.59, Sally's median incorrect response=17.75); whereas the lowest rates of correct and incorrect responses per minute were found in the cloze passages and the multiple-choice questions (cloze: Sally's median correct response=3, Sally's median incorrect response=0;

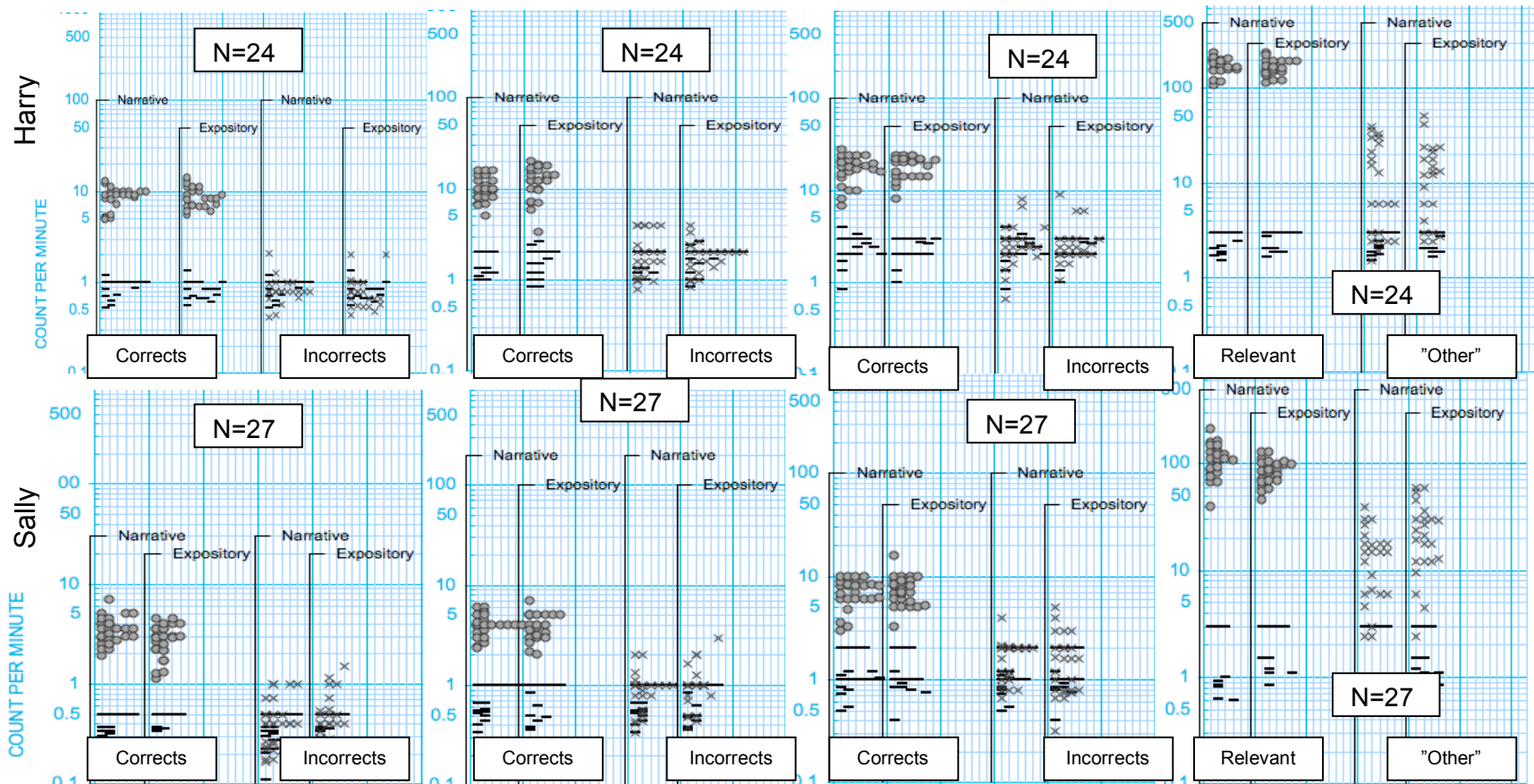*Figure 1.* Top graphs – Harry's frequency distributions of correct, relevant responses and incorrect responses for cloze passages (far left), multiple-choice questions (left middle), the sentence verification (right middle), and read and recall tests in the narrative or expository text conditions. Bottom graph - Lucy's frequency distributions of correct, relevant responses and incorrect responses.

multiple-choice questions: median correct response rate=3.8, median incorrect

response rate=1).

For Lucy (Figure 2), regardless of text condition, the graphs show that the

highest rates of both correct and incorrect responses per minute occurred in the read

and recall test (Lucy's median correct response=71.17, and Lucy's median incorrect

response=30). Whereas, the lowest rates of correct and incorrect responses per minute

were found in the cloze passages (Lucy's median correct response=2, and Lucy's

median incorrect response=.8).

<div align="center">Text Type Comparison</div>

Frequency Distributions. For Harry (top row, Figure 1), there was no difference

between the frequency distribution of correct responses in the narrative condition and

that of the expository condition for the cloze passages, sentence verification, or read

and recall tests. However, for the multiple-choice questions, Harry showed a slightly

lower frequency of correct responses per minute in the narrative text condition

(median=9.8, mode=8, and a range of 5 to 16) than he did in the expository condition

(median =12, mode=12, and a range of 3.3 to 20). There was no difference between

Harry's frequency distribution of incorrect responses in the narrative condition and that

of the expository condition for cloze passages, multiple-choice questions, or the

sentence verification test. However, during the read and recall test, Harry recalled a

lower frequency of "other" words per minute in the narrative condition (median=6,

mode=0 and a range of 0 to 39) than he did in the expository condition (median=12.56,

mode=6 and a range of 0 to 51).

For cloze passages (top left, Figure 1), the median rate of correct responses for

<div align="center">30</div>

the narrative and expository conditions were 9 and 8.3 respectively. The mode for the rate of correct responses for the narrative and expository conditions were 9 and 7 respectively. The rate of correct responses had a range of 5 to 13 in the narrative condition and a range of 5.5 to 14 in the expository condition. The median rate of incorrect responses for both the narrative and expository conditions was 0. The mode for the rate of incorrect responses for both the narrative and expository conditions was 0. The rate of incorrect responses had a range of 0 to 2.1 in the narrative condition and a range of 0 to 2 in the expository condition.

For multiple-choice questions (middle left, Figure 1), the median rate of correct responses for the narrative and expository conditions were 9.8 and 12 respectively. The mode for the rate of correct responses for the narrative and expository conditions were 8 and 12 respectively. The rate of correct responses had a range of 5 to 16 in the narrative condition and a range of 3.3 to 20 in the expository condition. The median rate of incorrect responses for the narrative and expository conditions were 1.7 and 2 respectively. The mode for the rate of incorrect responses for the narrative and expository conditions were 0 and 2 respectively. The rate of incorrect responses had a range of 0 to 4 in the narrative condition and a range of 0 to 4 in the expository condition.

For the sentence verification (middle right, Figure 1), the median rate of correct responses in the narrative and expository conditions were 18 and 20 respectively. The mode for the rate of correct responses for the narrative and expository conditions were 24 and 14 respectively. The rate of correct responses had a range of 6.7 to 28 in the narrative condition and a range of 8 to 24 in the expository condition. The median rate

31

of incorrect responses for the narrative and expository conditions were 2.4 and 2 respectively. The mode for the rate of incorrect responses in the narrative and expository conditions were 0 and 2 respectively. The rate of incorrect responses had a range of 0 to 8 in the narrative condition and a range of 0 to 9 in the expository condition.

For the recall test (far right, Figure 1), the median rate of correct responses per minute in the narrative and expository conditions were 169.7 and 159.7 respectively. The mode for the rate of correct responses in the narrative and expository conditions were 204 and 195 respectively. The rate of correct responses had a range of 108 to 240 in the narrative condition and a range of 114 to 240 in the expository condition. The median rate of incorrect responses in the narrative and expository conditions were 6 and 12.6 respectively. The mode for the rate of incorrect responses in the narrative and expository conditions were 0 and 6 respectively. The rate of incorrect responses had a range of 0 to 39 in the narrative condition and a range of 0 to 51 in the expository condition.

For Sally (bottom row, Figure 1), there was no difference between the frequency distribution of correct responses in the narrative condition and that of the expository condition for the cloze passage, multiple-choice questions or the sentence verification tests. The only difference was seen in the read and recall test where Sally recalled more relevant words per minute in the narrative condition (median=110.77, mode=66, and a range of 39 to 210) than she did in the expository condition (median=84, mode=84, and a range of 45 to 129). There was no difference between Sally's frequency distribution of incorrect responses in the narrative condition and that of the expository condition for the

32

cloze passage, multiple-choice questions, or the sentence verification test. However, during the read and recall test, Sally recalled fewer "other" words per minute in the narrative condition (median=15, mode=15 and a range of 0 to 39) than she did in the expository condition (median=21, mode=12 and a range of 0 to 60).

For cloze passages (far left, Figure 1), the median rate of correct responses for the narrative and expository conditions were 3.1 and 2.9 respectively. The mode for the rate of correct responses for the narrative and expository conditions were 3.5 and 3 respectively. The rate of correct responses had a range of 1.8 to 5 in the narrative condition and a range of 1.1 to 4.5 in the expository condition. The median rate of incorrect responses for the narrative and expository conditions were both 0. The mode for the rate of incorrect responses for the narrative and expository conditions were both 0. The rate of incorrect responses had a range of 0 to 1 in the narrative condition and a range of 0 to 1.5 in the expository condition.

For multiple-choice questions (middle left, Figure 1), the median rate of correct responses in the narrative and expository conditions were 4 and 3.6 respectively. The mode for the rate of correct responses in the narrative and expository conditions were 4 and 5 respectively. The rate of correct responses had a range of 2.3 to 6 in the narrative condition and a range of 1.6 to 7 in the expository condition. The median rate of incorrect responses in both the narrative and expository conditions was 1. The mode for the rate of incorrect responses for both the narrative and expository conditions was 1. The rate of incorrect responses had a range of 0 to 3 in the narrative condition and a range of 0 to 2 in the expository condition.

For the sentence verification (middle right, Figure 1), the median rate of correct responses for the narrative and expository conditions were 7.3 and 7.5 respectively. The mode for the rate of correct responses for the narrative and expository conditions were 6 and 8 respectively. The rate of correct responses had a range of 3 to 10 in the narrative condition and a range of 3 to 16 in the expository condition. The median rate of incorrect responses for the narrative and expository conditions were 2 and 1 respectively. The mode for the rate of incorrect responses for the narrative and expository conditions were 2 and 0 respectively. The rate of incorrect responses had a range of 0 to 4 in the narrative condition and a range of 0 to 5 in the expository condition.

For the recall test (far right, Figure 1), the median rate of correct responses for the narrative and expository conditions were 110.8 and 84 respectively. The mode for the rate of correct responses for the narrative and expository conditions were 66 and 84 respectively. The rate of correct responses had a range of 39 to 210 in the narrative condition and a range of 45 to 129 in the expository condition. The median rate of incorrect responses for the narrative and expository conditions were 15 and 21 respectively. The mode for the rate of incorrect responses for the narrative and expository conditions were 15 and 12 respectively. The rate of incorrect responses had a range of 0 to 39 in the narrative condition and a range of 0 to 60 in the expository condition.

For Lucy (Figure 2), there was no difference between the frequency distribution of correct responses in the narrative condition and that of the expository condition for the cloze passages, multiple-choice questions or the sentence verification tests.
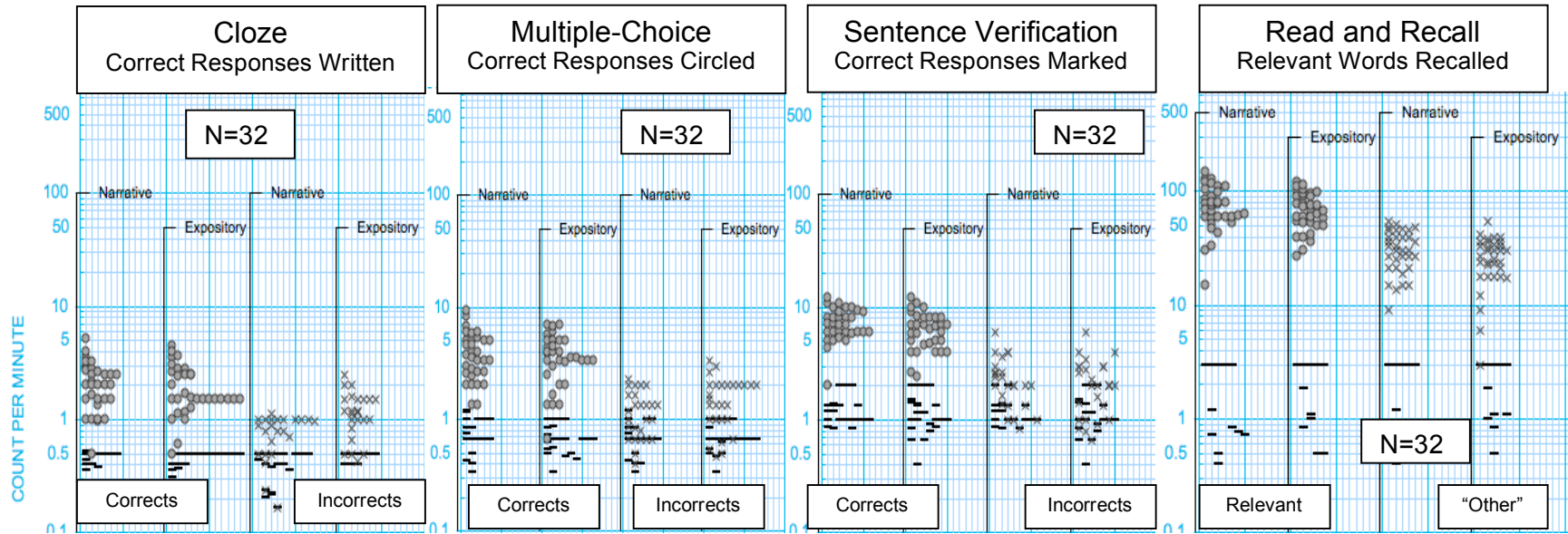
Figure 2. Frequency distributions of correct responses and incorrect responses for cloze passages (far left), multiple choice questions (left middle), and sentence verification (right middle) for Lucy in the narrative and expository text conditions. The frequency distributions of relevant words and "other" words recalled (far right) for Lucy across narrative and expository text conditions.

The only difference was seen in the read and recall test where Lucy recalled more relevant words per minute in the narrative condition (median=76.5, mode=60, and a range of 30 to 147) than she did in the expository condition (median=69.7, mode=90, and a range of 27 to 120). There was no difference between Lucy's frequency distribution of incorrect responses in the narrative condition and that of the expository condition for cloze passages, multiple-choice questions, or the sentence verification test. However, during the read and recall test, Lucy recalled fewer "other" words per minute in the narrative condition (median=30, mode=15 and a range of 9 to 54) than she did in the expository condition (median=26, mode=30 and a range of 3 to 54).

For cloze passages (far left, Figure 2), the median rate of correct responses for the narrative and expository conditions were 2 and 1.5 respectively. The mode for the rate of correct responses for the narrative and expository conditions were 2 and 1.5 respectively. The rate of correct responses had a range of .5 to 5.2 in the narrative condition and a range of .5 to 4.5 in the expository condition. The median rate of incorrect responses for the narrative and expository conditions were .7 and 1 respectively. The mode for the rate of incorrect responses for both the narrative and expository conditions was 0. The rate of incorrect responses had a range of 0 to 1.1 in the narrative condition and a range of 0 to 2.5 in the expository condition.

For multiple-choice questions (middle left, Figure 2), the median rate of correct responses for the narrative and expository conditions were 3.6 and 3.5 respectively. The mode for the rate of correct responses for the narrative and expository conditions were 5 and 3.3 respectively. The rate of correct responses had a range of 1.3 to 9.2 in the narrative condition and a range of .7 to 7 in the expository condition. The median

rate of incorrect responses for the narrative and expository conditions were 1 and 1.3 respectively. The mode for the rate of incorrect responses for the narrative and expository conditions were 0 and 2 respectively. The rate of incorrect responses had a range of 0 to 2.3 in the narrative condition and a range of 0 to 3.3 in the expository condition.

For the sentence verification (middle right, Figure 2), the median rate of correct responses for the narrative and expository conditions were 7.1 and 6.7 respectively. The mode for the rate of correct responses for the narrative and expository conditions were 6 and 4 respectively. The rate of correct responses had a range of 2 to 12 in the narrative condition and a range of 2.4 to 12 in the expository condition. The median rate of incorrect responses for the narrative and expository conditions were both 2. The mode for the rate of incorrect responses for the narrative and expository conditions were 0 and 2 respectively. The rate of incorrect responses had a range of 0 to 6 in both the narrative condition and the expository condition.

For the recall test (far right, Figure 2), the median rate of correct responses for the narrative and expository conditions were 76.5 and 69.7 respectively. The mode for the rate of correct responses for the narrative and expository conditions were 60 and 90 respectively. The rate of correct responses had a range of 30 to 147 in the narrative condition and a range of 27 to 120 in the expository condition. The median rate of incorrect responses for the narrative and expository conditions were 30 and 26 respectively. The mode for the rate of incorrect responses for the narrative and expository conditions were 15 and 30 respectively. The rate of incorrect responses had

a range of 9 to 54 in the narrative condition and a range of 3 to 54 in the expository

condition.

Celerations. Figures 3 and 4 show Harry, Sally, and Lucy's celerations of

responses per minute across the narrative and expository text conditions, for each test.

The celerations are based on graphs of daily individual performance for each reading

task. The individual graphs can be obtained from the experimenter by request (see a

sample in Appendix A). The left side of each graph displays the celerations of correct

responses or relevant words per minute and the right side of each graph presents the

celerations of incorrect responses or "other" words per minute.

For Harry (top row, Figure 3), regardless of the reading comprehension test

used, there was no consistent difference between Harry's celerations of correct and

incorrect responses per minute in the narrative and that of the expository condition.

For cloze passages (far left), the highest celerations of correct responses in the

narrative and the expository condition were both x1.1. The lowest celerations of correct

responses in the narrative and expository conditions were /1.1 and /1.5 respectively. For

incorrect responses, the highest celerations of incorrect responses in the narrative and

expository condition were x1.15 and x1.25 respectively. The lowest celerations of

incorrect responses per minute in the narrative and expository conditions were both x1.

For multiple-choice questions (middle left), the highest celerations of correct

responses in the narrative and the expository condition were both x1.2. The lowest

celerations of correct responses in the narrative and expository conditions were both

close to /1.1. For incorrect responses, the highest celerations of incorrect responses in

the narrative and expository condition were x1.2 and x1.1 respectively. The lowest
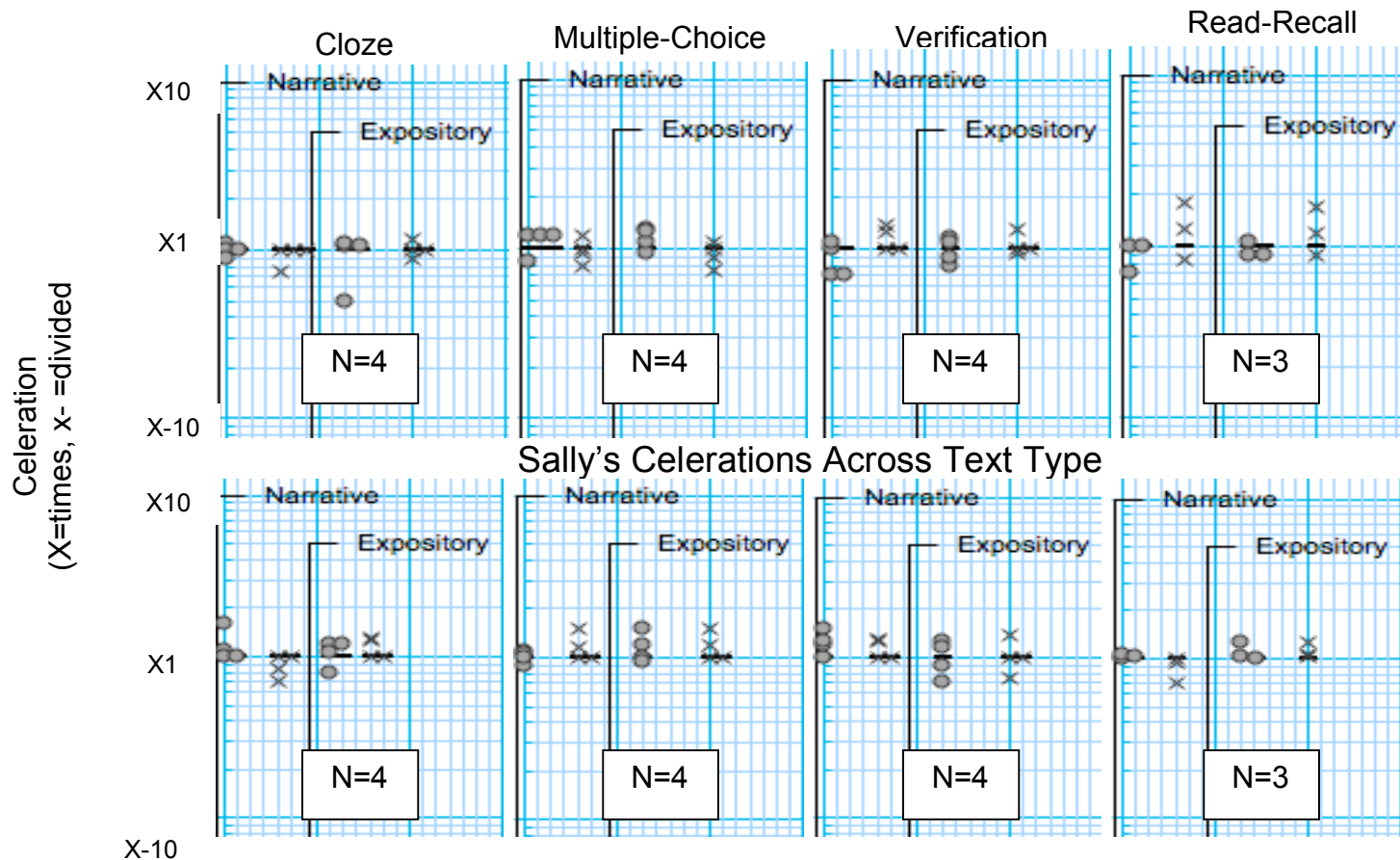
Figure 3. Top graphs - The celerations of correct and incorrect responses per minute for Harry across the narrative and expository text conditions for cloze passages (far left), multiple-choice questions (middle left), sentence verification (middle right), and read and recall (far right). Bottom graphs - The celerations of correct and incorrect responses per minute for Sally across the narrative and expository text conditions for the four reading comprehension tests. The solid dots are correct responses and the "X"s represent incorrect responses.

celerations of incorrect responses per minute in the narrative and expository conditions were both close to /1.2.

For the sentence verification (middle right), the highest celerations of correct responses in the narrative and the expository condition were x1 and x1.15 respectively. The lowest celerations of correct responses in the narrative and expository conditions were /1.3 and /1.2 respectively. For incorrect responses, the highest celerations of incorrect responses in the narrative and expository condition were x1.4 and x1.3 respectively. The lowest celerations of incorrect responses per minute in the narrative and expository conditions were both close to x1.

For the recall test (far right), the highest celerations of correct responses in the narrative and the expository condition were both x1.1. The lowest celerations of correct responses in the narrative and expository conditions were /1.1 and /1.5 respectively. For incorrect responses, the highest celerations of incorrect responses in the narrative and expository condition were x1.8 and x1.7 respectively. The lowest celerations of incorrect responses in the narrative and expository conditions x1.25 and /1.1 respectively.

For Sally (bottom row, Figure 3), regardless of the reading comprehension test used, there was no consistent difference between Sally's celerations of correct and incorrect responses per minute in the narrative condition and that of the expository condition.

For cloze passages (bottom left), the highest celerations of correct responses in the narrative and the expository condition were x1.6 and x1.2 respectively. The lowest celerations of correct responses in the narrative and expository conditions were x1 and /1.2 respectively. For incorrect responses, the highest celerations of incorrect responses

in the narrative and expository condition were x1 and x1.3 respectively. The lowest celerations of incorrect responses in the narrative and expository conditions were /1.3 and x1 respectively.

For multiple-choice questions (middle left), the highest celerations of correct responses in the narrative and the expository condition were x1.1 and x1.5 respectively. The lowest celerations of correct responses in the narrative and expository conditions were both close to x1. For incorrect responses, the highest celerations of incorrect responses in the narrative and expository condition were x1.2 and x1.1 respectively. The lowest celerations of incorrect responses in the narrative and expository conditions were both /1.2.

For the sentence verification (middle right), the highest celerations of correct responses in the narrative and the expository condition were x1.5 and x1.25 respectively. The lowest celerations of correct responses in the narrative and expository conditions were x1 and /1.3 respectively. For incorrect responses, the highest celerations of incorrect responses in the narrative and expository condition were x1.3 and x1.4 respectively. The lowest celerations of incorrect responses per minute in the narrative and expository conditions were x1 and /1.25 respectively.

For the read and recall test (far right), the highest celerations of relevant words recalled for narrative and expository were x1.05 and x1.25 respectively. The lowest celerations of relevant words recalled for the narrative and expository conditions were both x1. For "other" words recalled, the highest celerations of "other" words in the narrative and expository condition were x1 and x1.25 respectively. The lowest

celerations of "other" words in the narrative and expository conditions were /1.3 and x1 respectively.

For Lucy (Figure 4), regardless of the reading comprehension test used, there was no consistent difference between the celerations of correct and incorrect responses per minute in the narrative condition and that of the expository condition.

For cloze passages (far left), the highest celerations of correct responses in the narrative and the expository condition were both x1.15. The lowest celerations of correct responses in the narrative and expository conditions were x1 and /1.25 respectively. For incorrect responses, the highest celerations of incorrect responses in the narrative and expository condition were x1 and x1.3 respectively. The lowest celerations of incorrect responses in the narrative and expository conditions were /1.3 and x1 respectively.

For multiple-choice questions (middle left), the highest celerations of correct responses in the narrative and the expository condition were both close to x1. The lowest celerations of correct responses in the narrative and expository conditions were /1.3 and /1.1 respectively. For incorrect responses, the highest celerations of incorrect responses in the narrative and expository condition were both x1.2. The lowest celerations of incorrect responses in the narrative and expository conditions were /1.05 and /1.25 respectively.

For the sentence verification (middle right), the highest celerations of correct responses in the narrative and the expository condition were x1.05 and x1.2 respectively. The lowest celerations of correct responses in the narrative and expository conditions were /1.3 and x1.05 respectively. For incorrect responses, the highest celerations of incorrect responses in the narrative and expository condition were
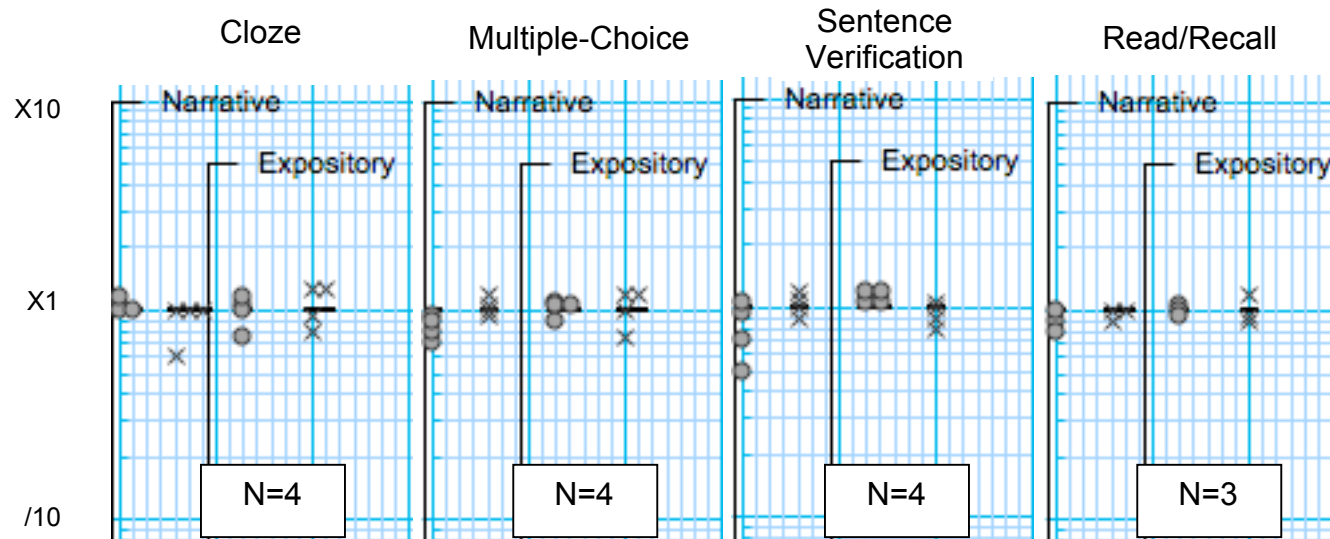
Figure 4. The celerations of correct and incorrect responses per minute for Lucy across narrative and expository text conditions for cloze passages (far left), multiple-choice questions (middle left), sentence verification (middle right) and read and recall (far right).

x1.2 and x1.05 respectively. The lowest celerations of incorrect responses per minute in the narrative and expository conditions were /1.1 and /1.2 respectively.

For the read and recall test (far right), the highest celerations of relevant words recalled for narrative and expository were both close to x1. The lowest celerations of relevant words recalled for the narrative and expository conditions were /1.2 and /1.05 respectively. For "other" words recalled, the highest celerations of "other" words in the narrative and expository condition were x1 and x1.2 respectively. The lowest celerations of "other" words in the narrative and expository conditions were both /1.1.

Time-Limit Condition (Time-limit or Duration)

Frequency Distributions. Figures 5, 6 and 7 show Harry, Sally, and Lucy's frequency distributions of correct and incorrect responses for the cloze passages (top left), multiple-choice questions (top right), sentence verification (bottom left) and read and recall tests (bottom right) across duration and time limit conditions. For each task, the correct responses or relevant words per minute are presented on the left-hand side of the graph, while the incorrect responses or "other" words per minute are shown on the right-hand side of the graph. The responses are displayed as a count per minute and show horizontally, how frequently a specific score occurred in each condition. The "N"s on the graph tell the number of sessions that occurred in each condition.

For Harry (Figure 5), regardless of the reading comprehension test used, there was no consistent difference between the frequency of correct or incorrect responses in the duration and time limit testing conditions.

For cloze passages, for the first duration condition and the first time-limit condition Harry's frequency distributions of correct responses had medians of 8.3 and 9
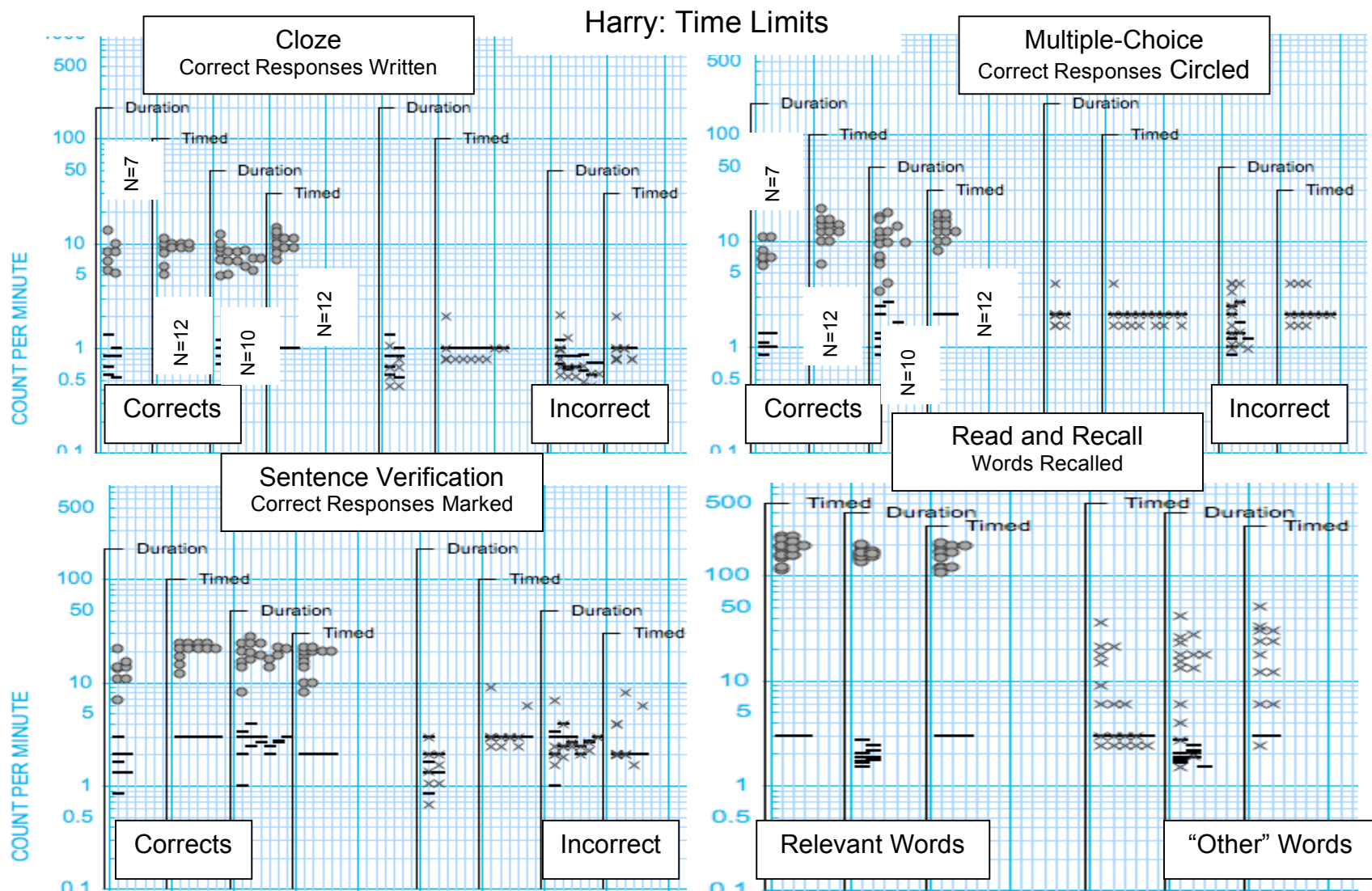
44

Figure 5. Harry's frequency distributions of correct, relevant responses and incorrect responses for cloze passages (top left), multiple-choice questions (top right), the sentence verification (bottom left), and read and recall tests (bottom right) in the duration timing and time-limit timing conditions.

respectively. For the second duration condition and the second time-limit condition the median rate of correct responses were 7.1, and 10.5 respectively. The first duration and time-limit testing conditions had modes of 8.3 and 10 respectively. The second duration and time limit conditions had modes of 8.3 and 9 respectively. The first duration and time limit conditions had ranges of 5.2 to 13.3 and 5 to 11 respectively. The second duration and time limit conditions had ranges of 4.8 to 12 and 7 to 14 respectively. For incorrect responses, for all the duration and time limit conditions, Harry's frequency distributions of incorrect responses per minute had medians and modes of 0. In the first duration and time limit conditions, the rate of incorrect responses had ranges of 0 to 0 and 0 to 2 respectively. The second duration and time limit conditions had ranges of 0 to 2.1 and 0 to 2 respectively.

For the multiple-choice questions (top right), for the first duration condition and the first time-limit condition Harry's frequency distributions of correct responses had medians of 7.5 and 13 respectively. For the second duration condition and the second time-limit condition the median rate of correct responses were 10.6 and 13 respectively. The first duration and time-limit testing conditions had modes of 8 and 14 respectively. The second duration and time limit conditions had modes of 10.6 and 16 respectively. The first duration and time limit condition had ranges of 5.8 to 12 and 8 to 16 respectively. The second duration and time limit condition had ranges of 3.3 to 17.5 and 8 to 18 respectively. For incorrect responses, for the first duration and time limit conditions, the median rates of incorrect responses per minute were .9 and 1 respectively. For the second duration and time limit conditions, the median rates of incorrect responses per minute were 1.2 and 2 respectively. For the first duration and

time limit conditions, the mode rates of incorrect responses per minute were 0 and 2 respectively. For the second duration and time limit conditions, the mode rates of incorrect responses per minute were 0 and 2 respectively. In the first duration and time limit condition, the rate of incorrect responses had ranges of 0 to 2.2 and 0 to 2 respectively. The second duration and time limit condition had ranges of 0 to 3.3 and 0 to 4 respectively.

For the sentence verification (middle right), for the first duration condition and the first time limit condition Harry's frequency distributions of correct responses had medians of 14 and 21 respectively. For the second duration condition and the second time-limit condition the median rate of correct responses were 19 and 21 respectively. The first duration and time-limit testing conditions had modes of 14 and 21 respectively. The second duration and time limit conditions had modes of 24 and 20 respectively. The first duration and time limit condition had ranges of 6.7 to 17.1 and 12 to 24 respectively. The second duration and time limit condition had ranges of 8 to 24 and 8 to 22 respectively. For incorrect responses, for the first duration and time limit conditions, the median rates of incorrect responses per minute were 0 and 3 respectively. For the second duration and time limit conditions, the median rates of incorrect responses per minute were both 2. For the first duration and time limit conditions, the mode rates of incorrect responses per minute were 0 and 3 respectively. For the second duration and time limit conditions, the mode rates of incorrect responses per minute were 2 and 0 respectively. In the first duration and time limit condition, the rate of incorrect responses had ranges of 0 to 2 and 0 to 9 respectively. The second duration and time limit condition had ranges of 0 to 3 and 0 to 8 respectively.

For the recall test (bottom right), for the first time-limit condition, the first duration condition and the reversal to a time limit testing condition Harry's frequency distributions of correct responses had medians of 175.5, 155.1 and 165 respectively. The first time-limit testing condition, the first duration testing condition and the reversal to a time limit testing condition had modes of 195, 171, and 195 respectively. The first time limit and duration conditions had ranges of 114 to 240 and 135 to 171.8 respectively. The reversal to a time-limit testing condition had a range of 108 to 204. For incorrect responses, for the first time-limit testing condition, the first duration testing condition and the reversal to a time-limit testing condition, the median rates of incorrect responses per minute were 6, 13.1 and 18 respectively. For the first time-limit testing condition, the first duration condition and the reversal to a time-limit testing condition, the mode rates of incorrect responses per minute were 3, 13.1 and 24 respectively. In the first time-limit testing condition and the duration testing condition, the rate of incorrect responses had ranges of 0 to 36 and 0 to 42 respectively. The reversal to a time-limit testing condition had a range of 0 to 51.

For Sally (Figure 6), there was no difference in the frequency of correct or incorrect responses per minute between the duration and time-limit testing conditions in the cloze passages, the multiple-choice questions and recall tests. However, a difference was found between testing conditions in the sentence verification test. Sally showed a higher frequency of both correct and incorrect responses in the time-limit condition than she did in the duration condition.

For cloze passages (top left), for the first duration condition and the first time-limit condition Sally's frequency distributions of correct responses had medians of 2.3 and
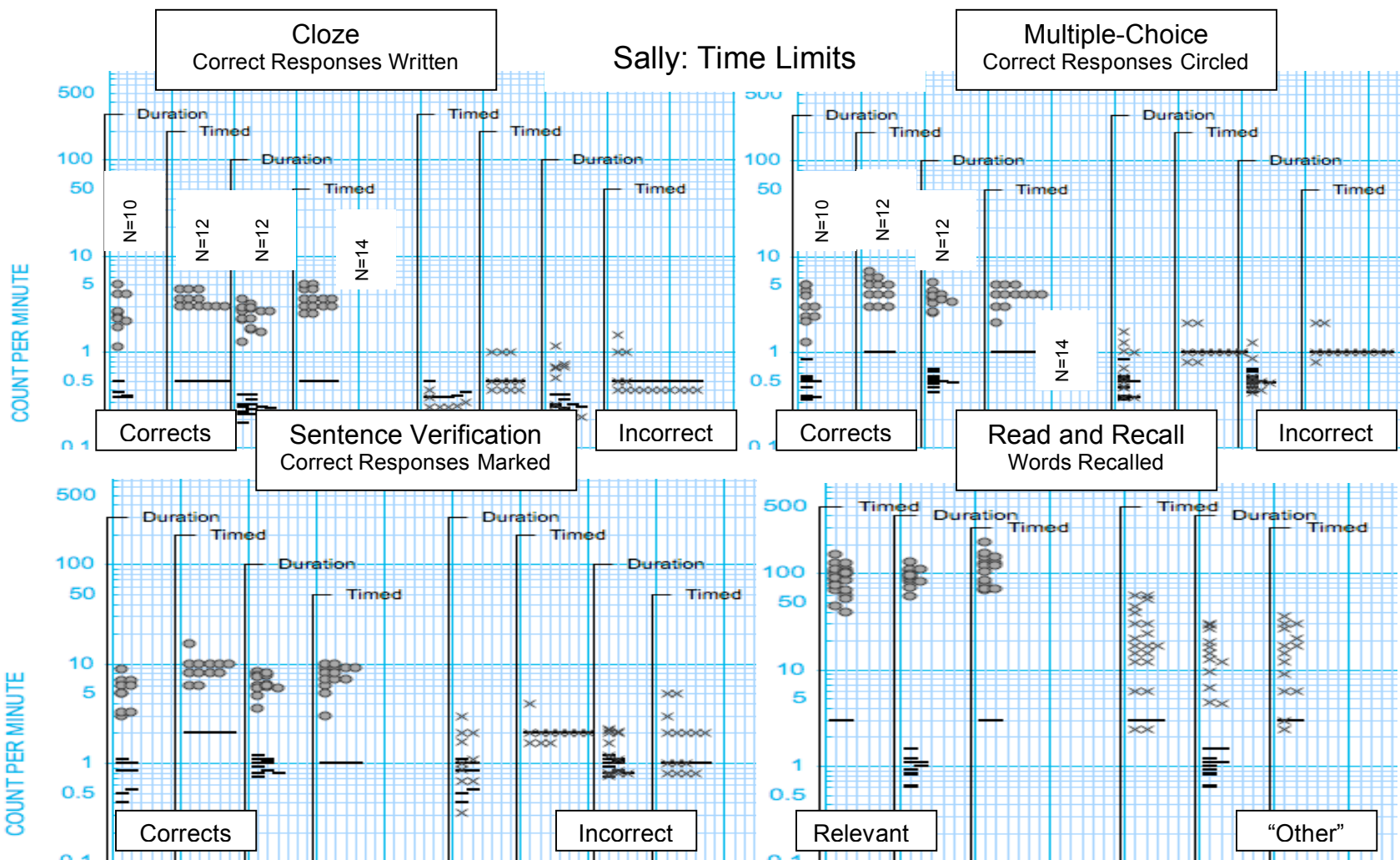
Figure 6. Sally's frequency distributions of correct, relevant responses and incorrect responses for cloze passages (top left), multiple-choice questions (top right), sentences verification (bottom left), and read and recall tests (bottom right) in the duration timing and time-limit timing conditions.

3.5 respectively. For the second duration condition and the second time-limit condition the median rate of correct responses were 2.6 and 3.5 respectively. The first duration and time-limit testing conditions had modes of 4 and 3 respectively. The second duration and time limit conditions had modes of 2.8 and 3 respectively. The first duration and time limit conditions had ranges of 1.1 to 4 and 3 to 5 respectively. The second duration and time limit conditions had ranges of 1.3 to 3.5 and 2.5 to 5 respectively. For incorrect responses, for all the duration and time limit conditions, Sally's frequency distributions of incorrect responses per minute had medians and modes of 0. In the first duration and time limit conditions, the rate of incorrect responses had ranges of 0 to 3 and 0 to 1 respectively. The second duration and time limit conditions had ranges of 0 to 1.1 and 0 to 1.5 respectively.

For the multiple-choice questions (top right), for the first duration condition and the first time-limit condition Sally's frequency distributions of correct responses had medians of 2.7 and 5 respectively. For the second duration condition and the second time-limit condition the median rate of correct responses were 3.6 and 4 respectively. The first duration and time-limit testing conditions had modes of none and 3 respectively. The second duration and time limit conditions had modes of 3.5 and 4 respectively. The first duration and time limit condition had ranges of 1.6 to 5 and 3 to 7 respectively. The second duration and time limit condition had ranges of 2.6 to 5.3 and 2 to 5 respectively. For incorrect responses, for the first duration and time limit conditions, the median rates of incorrect responses per minute were .9 and 1 respectively. For the second duration and time limit conditions, the median rates of incorrect responses per minute were .5 and 1 respectively. For the first duration and

time limit conditions, the mode rates of incorrect responses per minute were .3 and 1 respectively. For the second duration and time limit conditions, the mode rates of incorrect responses per minute were 0 and 1 respectively. In the first duration and time limit condition, the rate of incorrect responses had ranges of 0 to 1.7 and 0 to 2 respectively. The second duration and time limit condition had ranges of 0 to 1.25 and 0 to 3 respectively.

For the sentence verification test (bottom left), for the first duration condition and the first time limit condition Sally's frequency distributions of correct responses had medians of 5.5 and 8 respectively. For the second duration condition and the second time-limit condition the median rate of correct responses were 6.3 and 8 respectively. The first duration and time-limit testing conditions had modes of 6.7 and 10 respectively. The second duration and time limit conditions had modes of 8 and 9 respectively. The first duration and time limit condition had ranges of 3 to 8.7 and 6 to 10 respectively. The second duration and time limit condition had ranges of 4.3 to 8.4 and 3 to 10 respectively. For incorrect responses, for the first duration and time limit conditions, the median rates of incorrect responses per minute were 1 and 2 respectively. For the second duration and time limit conditions, the median rates of incorrect responses per minute were 1.1 and 1.5. For the first duration and time limit conditions, the mode rates of incorrect responses per minute were 0 and 2 respectively. For the second duration and time limit conditions, the mode rates of incorrect responses per minute were both 0. In the first duration and time limit condition, the rate of incorrect responses had ranges of 0 to 3 and 0 to 4 respectively. The second duration and time limit condition had ranges of 0 to 2 and 0 to 5 respectively.

For the recall test (bottom right), for the first time-limit condition, the first duration condition and the reversal to a time limit testing condition Sally's frequency distributions of correct responses had medians of 94.5, 93.8 and 120 respectively. The first time-limit testing condition, the first duration testing condition and the reversal to a time limit testing condition had modes of 84, 81 and 120 respectively. The first time limit and duration conditions had ranges of 18 to 159 and 58.3 to 129.8 respectively. The reversal to a time-limit testing condition had a range of 66 to 210. For incorrect responses, for the first time-limit testing condition, the first duration testing condition and the reversal to a time-limit testing condition, the median rates of incorrect responses per minute were 18, 14 and 16.5 respectively. For the first time-limit testing condition, the first duration condition and the reversal to a time-limit testing condition, the mode rates of incorrect responses per minute were 18, none and 6 respectively. In the first time-limit testing condition and the duration testing condition, the rate of incorrect responses had ranges of 0 to 60 and 4.5 to 30 respectively. The reversal to a time-limit testing condition had a range of 0 to 36.

For Lucy (Figure 7), there was no difference in the frequency of correct or incorrect responses per minute between the duration and time-limit testing conditions in the cloze passages, multiple-choice questions or sentence verification. The only difference was found between duration and time limit testing conditions in the read and recall test. Lucy showed a higher frequency of both correct and incorrect responses in the time-limit condition than she did in the duration condition.

For cloze passages (top right), for the first duration condition and the first time-limit condition Lucy's frequency distributions of correct responses had medians of 3 and
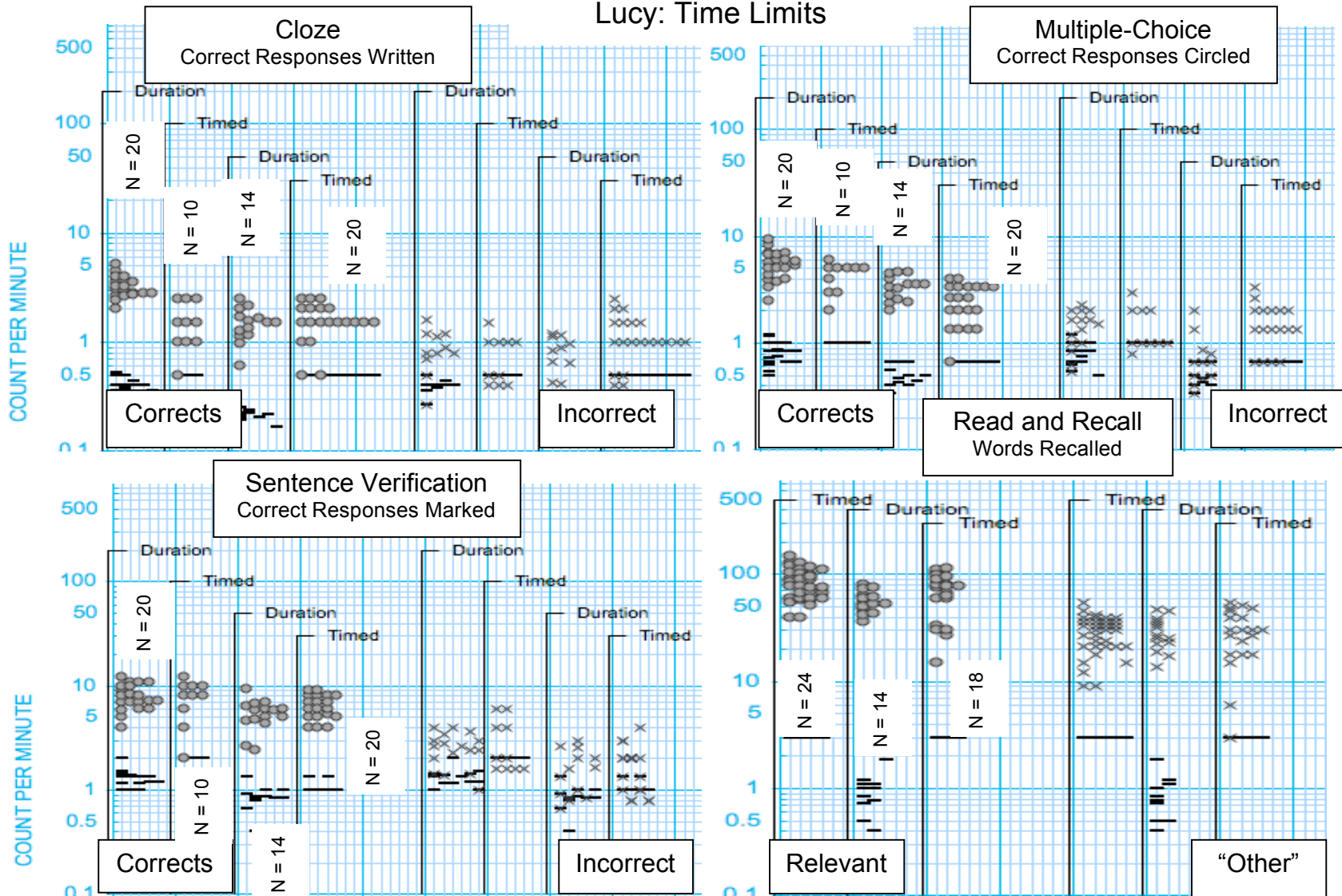
Figure 7. Lucy's frequency distributions of responses for cloze passages (top left), multiple-choice questions (top right), sentence verification (bottom left), and recall tests (bottom right) in the duration and time limit condition.

1.5 respectively. For the second duration condition and the second time-limit condition the median rate of correct responses were both 1.5. The first duration and time-limit testing conditions had modes of 2.8 and 2.5 respectively. The second duration and time limit conditions both had modes of 1.5. The first duration and time limit conditions had ranges of 2 to 5.2 and .5 to 2.5 respectively. The second duration and time limit conditions had ranges of .6 to 2.5 and .5 to 2.5 respectively. For incorrect responses, the first duration condition and the first time-limit condition, Lucy's frequency distributions of incorrect responses per minute had medians of .4 and .8 respectively. In the second duration condition and the second time-limit condition, the median rates of incorrect responses per minute were .7 and 1 respectively. In the first duration condition and the first time-limit condition, Lucy's frequency distributions of incorrect responses per minute had modes of 0 and 1 respectively. In the second duration condition and the second time-limit condition, the mode rates of incorrect responses per minute were both 1. In the first duration and time limit conditions, the rate of incorrect responses had ranges of 0 to 1.6 and 0 to 1.5 respectively. The second duration and time limit conditions had ranges of 0 to 1.2 and 0 to 2.5 respectively.

For the multiple-choice questions (top right), for the first duration condition and the first time-limit condition Lucy's frequency distributions of correct responses had medians of 5.5 and 5 respectively. For the second duration condition and the second time-limit condition the median rate of correct responses were 3.4 and 2.3 respectively. The first duration and time-limit testing conditions both had modes of 5. The second duration and time limit conditions had modes of 3.5 and 3.3 respectively. The first duration and time limit condition had ranges of 2.5 to 9.3 and 2 to 6 respectively. The

second duration and time limit condition had ranges of 1.7 to 5.7 and .7 to 4 respectively. For incorrect responses, for the first duration and time limit conditions, the median rates of incorrect responses per minute were 1.1 and 1 respectively. For the second duration and time limit conditions, the median rates of incorrect responses per minute were .7 and 1.3 respectively. For the first duration and time limit conditions, the mode rates of incorrect responses per minute were 0 and 1 respectively. For the second duration and time limit conditions, the mode rates of incorrect responses per minute were 0 to 1.3 respectively. In the first duration and time limit condition, the rate of incorrect responses had ranges of 0 to 2.3 and 0 to 3 respectively. The second duration and time limit condition had ranges of 0 to 2 and 0 to 3.3 respectively.

For the sentence verification test (bottom left), for the first duration condition and the first time limit condition Lucy's frequency distributions of correct responses had medians of 7.6 and 8 respectively. For the second duration condition and the second time-limit condition the median rate of correct responses were 5.6 and 6 respectively. The first duration and time-limit testing conditions had modes of 10.7 and 10 respectively. The second duration and time limit conditions had modes of 5 and 6 respectively. The first duration and time limit condition had ranges of 4 to 12 and 2 to 12 respectively. The second duration and time limit condition had ranges of 2.4 to 9.3 and 4 to 9 respectively. For incorrect responses, for the first duration and time limit conditions, the median rates of incorrect responses per minute were 2.4 and 2 respectively. For the second duration and time limit conditions, the median rates of incorrect responses per minute were 1.3 and 2. For the first duration and time limit conditions, the mode rates of incorrect responses per minute were both 0. For the

second duration and time limit conditions, the mode rates of incorrect responses per minute were 1.3 and 2 respectively. In the first duration and time limit condition, the rate of incorrect responses had ranges of 0 to 4 and 0 to 6 respectively. The second duration and time limit condition had ranges of 0 to 3 and 0 to 4 respectively.

For the recall test (bottom right), for the first time-limit condition, the first duration condition and the reversal to a time limit testing condition Lucy's frequency distributions of correct responses had medians of 81, 52 and 78 respectively. The first time-limit testing condition, the first duration testing condition and the reversal to a time limit testing condition had modes of 60, 52 and 60 respectively. The first time limit and duration conditions had ranges of 39 to 147 and 36 to 79 respectively. The reversal to a time-limit testing condition had a range of 15 to 114. For incorrect responses, for the first time-limit testing condition, the first duration testing condition and the reversal to a time-limit testing condition, the median rates of incorrect responses per minute were 30, 26 and 30 respectively. For the first time-limit testing condition, the first duration condition and the reversal to a time-limit testing condition, the mode rates of incorrect responses per minute were 21, none and then, 30 respectively. In the first time-limit testing condition and the duration testing condition, the rate of incorrect responses had ranges of 9 to 54 and 13.6 to 46.4 respectively. The reversal to a time-limit testing condition had a range of 3 to 54.

Celerations. Figure 8, 9 and 10 show Harry, Sally, and Lucy's celerations of responses per minute across the time limit and duration conditions, for each task. The celerations are based on graphs of daily individual performance for each reading task. The left side of each graph displays the celerations of correct responses or relevant

words per minute and the right side of each graph presents the celerations of incorrect responses or "other" words per minute. The "N"s on the graph tell the number of phases for each condition.

For Harry (Figure 8), regardless of the reading comprehension test used, there was no consistent difference in Harry's celerations of correct and incorrect responses per minute between the duration test condition and the time-limit test condition.

For cloze passages (far left), the highest celerations of correct responses in the duration and the time-limit condition were x1.05 and x1.1 respectively and the lowest celerations were /1.5 and /1.1 respectively. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing conditions were x1 and x1.15 respectively and the lowest celerations were /1.25 and /1.1 respectively.

For multiple-choice questions (middle left), the highest celerations of correct responses in the duration and the time-limit condition were x1.3 and x1.25 respectively and the lowest celerations were /1.05 and /1.15 respectively. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing conditions were x1.2 and x1.1 respectively and the lowest celerations were /1.25 and /1.1 respectively.

For the sentence verification test (middle right), the highest celerations of correct responses in the duration and the time-limit condition were x1.3 and x1.1 respectively and the lowest celerations were x1.1 and /1.3 respectively. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing conditions were x1.25 and x1.4 respectively and the lowest celerations were /1.05 and x1 respectively.
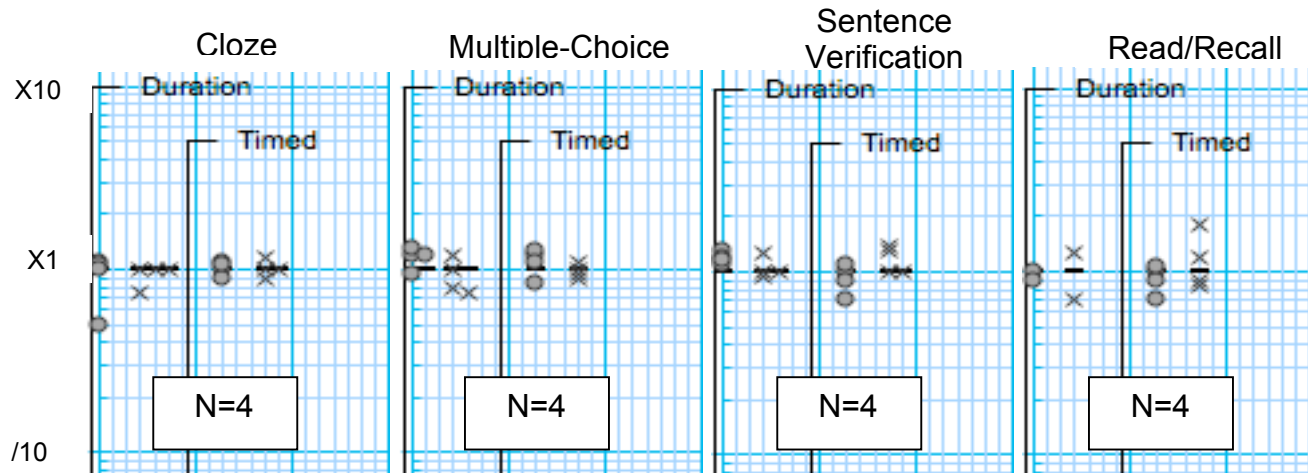
Figure 8. The celerations of correct and incorrect responses per minute for Harry across the duration and time limit test conditions for cloze passages (far left), multiple-choice questions (middle left), sentence verification (middle right), and read and recall (far right).

For the recall test (far right), the highest celerations of correct responses in the duration and the time-limit condition were x1.05 and x1 respectively and the lowest celerations were /1.3 and /1.1. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing condition were both x1.25 and the lowest celerations were /1.15 and /1.3 respectively.

For Sally (Figure 9), regardless of the reading comprehension test used, there was no consistent difference in Sally's celerations of correct and incorrect responses per minute between the duration test condition and the time-limit test condition.

For cloze passages (far left), the highest celerations of correct responses in the duration and the time-limit condition were x1.6 and x1.2 respectively and the lowest celerations were /1.2 and x1. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing conditions were x1.3 and x1.25 respectively and the lowest celerations were /1.3 and /1.15.

For multiple-choice questions (middle left), the highest celerations of correct responses in the duration and the time-limit condition were x1.5 and x1.2 respectively and the lowest celerations were /1.05 and /1.1. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing conditions were x1.5 and x1.2 respectively and the lowest celerations were both x1.

For the sentence verification test (middle right), the highest celerations of correct responses in the duration and the time-limit condition were x1.5 and x1.25 respectively and the lowest celerations were /1.1 and /1.3. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing conditions were x1.3 and x1.4 respectively and the lowest celerations were /1.25 and x1.
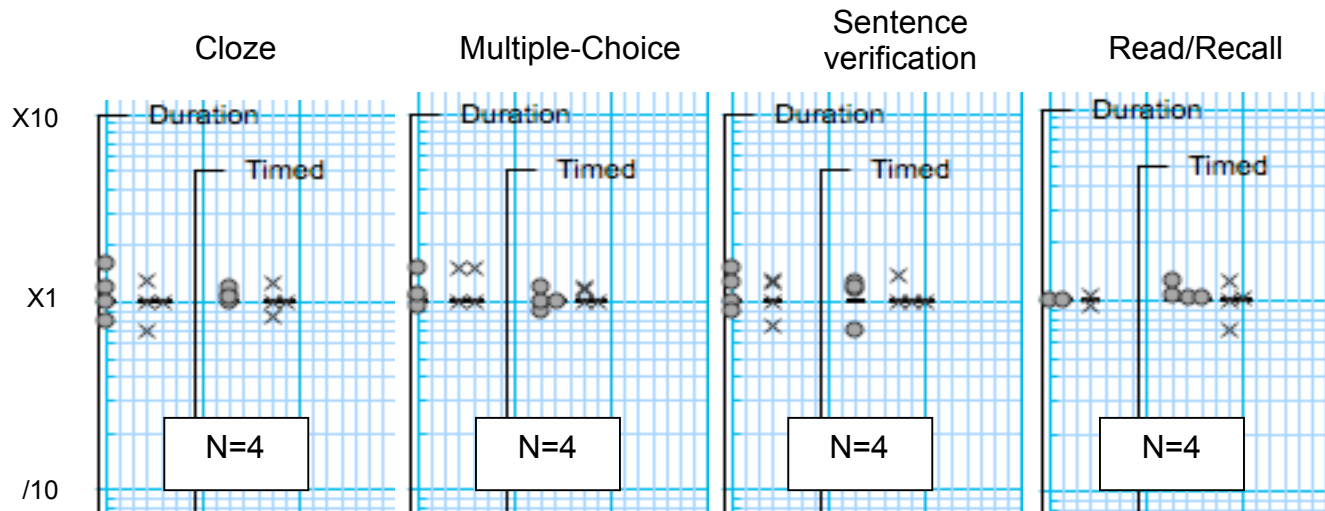
Figure 9. The celerations of correct and incorrect responses per minute for Sally across the duration and time limit test conditions for cloze passages (far left), multiple-choice questions (middle left), sentence verification (middle right), and read and recall (far right).

For the recall test (far right), the highest celerations of correct responses in the duration and the time-limit condition were x1 and x1.25 respectively and the lowest celerations were both x1. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing condition were x1.05 and x1.25 respectively and the lowest celerations were /1.05 and /1.3.

For Lucy (Figure 10), regardless of the reading comprehension test used, there was no consistent difference in Lucy's celerations of correct and incorrect responses per minute between the duration test condition and the time-limit test condition.

For cloze passages (far left), the highest celerations of correct responses in the duration and the time-limit condition were x1.15 and x1 respectively and the lowest celerations were x1 and /1.25. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing conditions were both x1.25 and the lowest celerations were /1.4 and /1.2.

For multiple-choice questions (middle left), the highest celerations of correct responses in the duration and the time-limit condition were both x1.1 and the lowest celerations were /1.1 and /1.3 respectively. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing conditions were both x1.2 and the lowest celerations were /1.25 and x1 respectively.

For the sentence verification test (middle right), the highest celerations of correct responses in the duration and the time-limit condition were x1.2 and x1.5 and the lowest celerations were /1.5 and /1.3 respectively. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing conditions were x1.1 and x1.2 respectively and the lowest celerations were /1.1 and /1.2 respectively.
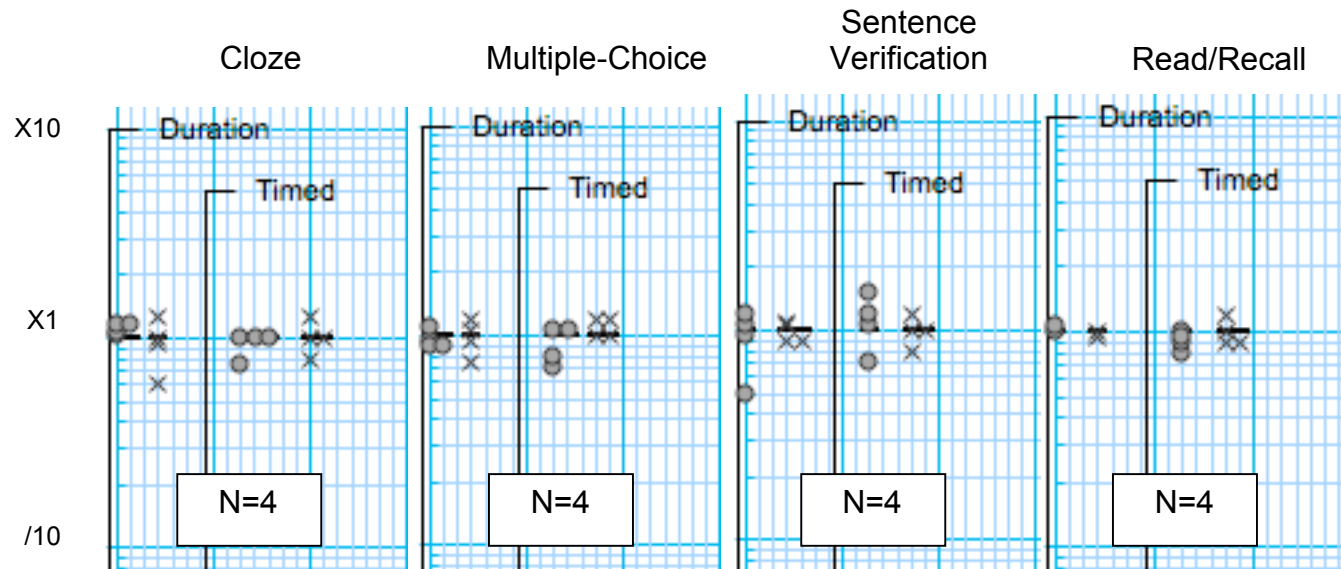
Figure 10. The celerations of correct and incorrect responses per minute for Lucy across duration and time-limit testing conditions for cloze passages (far left), multiple-choice questions (middle left), sentence verification (middle right), and read and recall (far right).

For the recall test (far right), the highest celerations of correct responses in the duration and the time-limit condition were x1.05 and x1 respectively and the lowest celerations were x1 and /1.2 respectively. For incorrect responses, the highest celerations of incorrect responses in the duration and time limit testing condition were x1 and x1.2 respectively and the lowest celerations were /1.05 and /1.1 respectively.

First 20 Seconds

Figure 11 shows a comparison of the daily frequency and celeration of the number of relevant and "other" words recalled per minute for the duration timing and for the first 20 seconds of the same duration timing for Harry (top graphs), Sally (middle graphs) and Lucy (bottom graphs). The graphs are presented vertically by participant and horizontally separated into the narrative and expository conditions. Across the x-axis are days and along the y-axis are responses per minute.

For Harry (top row, Figure 11), there was no difference between the daily frequencies or the celerations of relevant words or "other" words recalled during the duration timing and those found in the first 20 seconds of the recall timing in the narrative condition. The median frequency of relevant words recalled in the narrative condition for the duration timing and the first 20 seconds were both 168 relevant words. The median frequency of "other" words recalled in the narrative condition for the duration timing and the first 20 seconds were 6 and 24 respectively. The celerations of relevant and "other" words recalled in the narrative condition for the duration timing and the first 20 seconds were all x1. For Harry, during the expository condition, there was no difference between the daily frequencies or the celerations of relevant words recalled during the duration timing and those found in the first 20 seconds of the recall timing.

Figure 11. Top: Harry's number of relevant words recalled (solid dots) and "other" words recalled ("x"s) for the entire duration timing and the first 20 seconds of the duration timing across narrative and expository text conditions. Middle: Sally's number of relevant words recalled (solid dots) and "other" words recalled ("x"s). Bottom: Lucy's number of relevant words recalled (solid dots) and "other" words recalled ("x"s).

The median frequency of relevant words recalled in the expository condition for the duration timing and the first 20 seconds were 156 and 162 relevant words per minute respectively. The frequency of "other" words recalled was higher in the duration condition than it was in the first 20 seconds of the expository condition. The median frequency of "other" words recalled in the expository condition for the duration timing and the first 20 seconds were 13 and 6 "other" words per minute respectively. In the expository condition, there was no difference between the celeration of relevant words recalled in the duration condition and those of the first 20 seconds. In both conditions, the celerations were x1. However, the celerations of "other" words recalled in the expository condition were higher in the duration condition than they were in the first 20 seconds. The celerations of "other" words recalled in the duration condition and the first 20 seconds were x1.5 and x1 respectively.

For Sally (Middle Row, Figure 11), in the narrative condition, there was no difference between the daily frequencies or the celerations of relevant words recalled during the duration timing and thos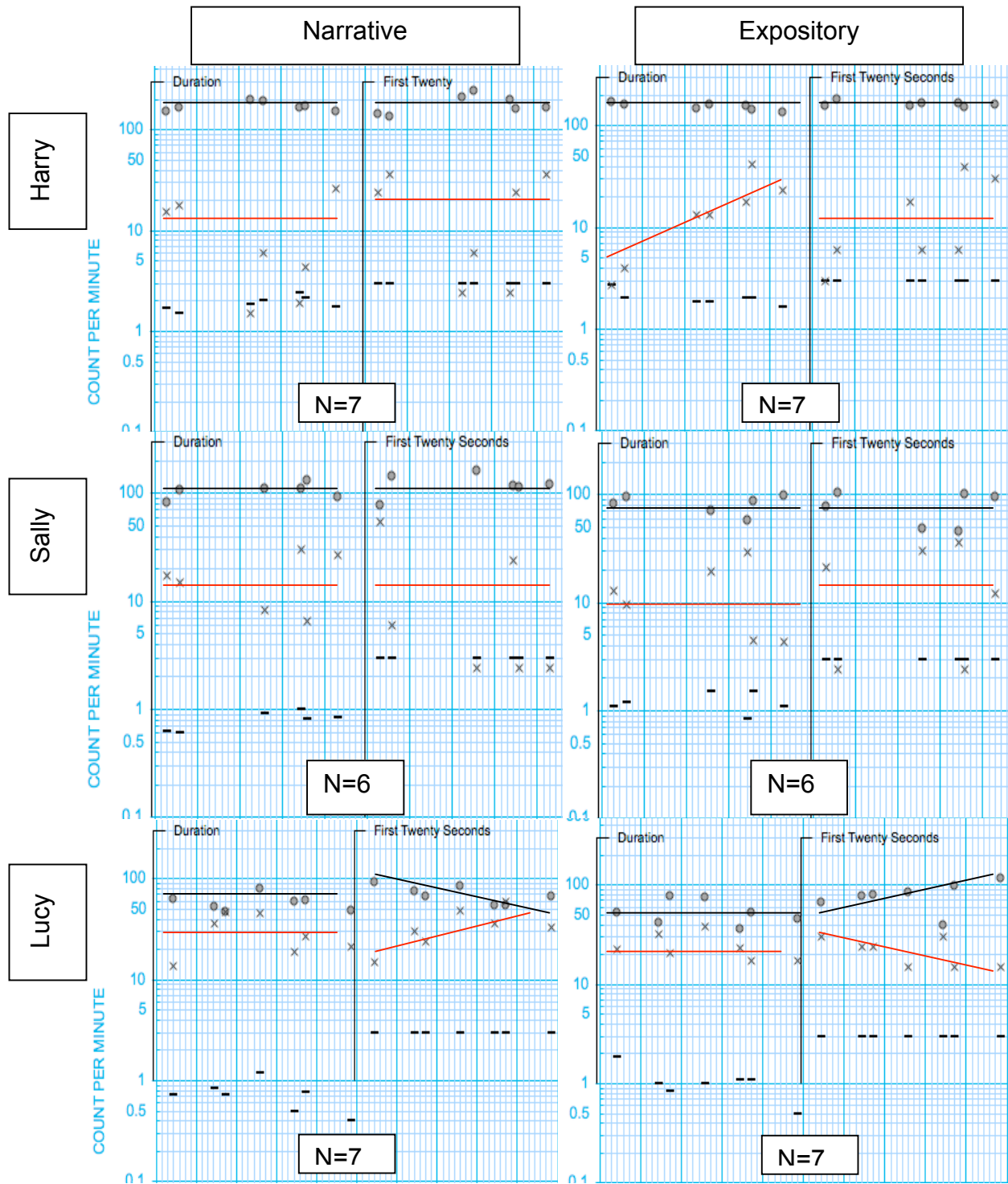e found in the first 20 seconds of the recall timing. The median frequency of relevant words recalled in the narrative condition for the duration timing and the first 20 seconds were 109 and 118 relevant words respectively. In the narrative condition, there was a higher frequency of "other" words recalled in the duration condition than was recalled during the first 20 seconds. The median frequency of "other" words recalled in the narrative condition for the duration timing and the first 20 seconds were 16 and 3 respectively. There was no difference between the celerations of "other" words recalled in the duration and first 20 seconds conditions. The celerations of relevant and "other" words recalled in the narrative condition for the duration timing

and the first 20 seconds were all x1. For Sally, in the expository condition, there was no difference between the daily frequencies or the celerations of relevant words or "other" words recalled during the duration timing and those found in the first 20 seconds of the recall timing. The median frequency of relevant words recalled in the expository condition for the duration timing and the first 20 seconds were 84 and 87 relevant words respectively. The median frequency of "other" words recalled in the expository condition for the duration timing and the first 20 seconds were 11 and 17 respectively. The celerations of relevant and "other" words recalled in the narrative condition for the duration timing and the first 20 seconds were all x1.

For Lucy (bottom row, Figure 11), during the narrative condition, there was no difference between the daily frequencies of relevant or "other" words recalled during the duration timing and those found in the first 20 seconds of the recall timing. The median frequency of relevant words recalled in the narrative condition for the duration timing and the first 20 seconds were 60 and 66 relevant words per minute respectively. The median frequency of "other" words recalled in the narrative condition for the duration timing and the first 20 seconds were 27 and 33 "other" words per minute respectively. In the narrative condition, there was no difference between the celeration of relevant words recalled in the duration condition and those of the first 20 seconds. In both conditions, the celerations were x1. However, the celerations of "other" words recalled in the narrative condition were more stable in the duration condition than they were in the first 20 seconds. The celerations of "other" words recalled in the duration condition and the first 20 seconds were x1 and x1.25 respectively. For Lucy, in the expository condition, both the daily frequencies and the celerations of relevant words recalled per

66

minute were lower and more stable during the duration condition than the frequencies and celeration in the first 20 seconds. The median frequency of relevant words recalled in the expository condition for the duration timing and the first 20 seconds were 52 and 81 relevant words per minute respectively. For "other" words, there was no difference between the frequency of "other" words recalled for the entire duration timing and in the first 20 seconds. The median frequency of "other" words recalled in the narrative condition for the duration timing and the first 20 seconds were 23 and 24 "other" words per minute respectively. The celeration of "other" words recalled in the expository condition was more stable in the entire duration than it was in the first 20 seconds. The celeration of "other" words recalled for the entire duration and the first 20 seconds were x1 and /1.25 respectively.

DISCUSSION

One purpose of the study was to examine whether text format (narrative or expository) has an effect on reading comprehension assessment scores. The results of the study showed no difference in the frequency distributions of correct responses per minute in the narrative condition, versus that of the expository condition, as measured by the cloze passages, multiple-choice questions or the sentence verification tests (see Figures 1 and 2). None of the participants showed a consistent difference in celeration under either narrative or expository conditions, using any of the selected reading comprehension tests (see Figures 3 and 4). Interestingly, the recall test appears to be the only assessment sensitive to text type. Both Sally and Lucy recalled more relevant words per minute in the narrative condition than in the expository condition, but Harry did not (See Figures 1 and 2).

The discrepancy between performance on the recall test in the narrative condition and that of the expository condition should not be ignored. Since the differences between narrative and expository texts were seen only in the recall test, it tells us something about the relationship between the assessments. Without a uniform effect, our results indicate that different forms of assessment may detect different component behaviors. It is possible that by measuring widely we were able to capture differences with the recall test not detected in the other assessments. The discrepancies in the recall test between narrative and expository conditions reveal the possible influences of text type and show us the sensitivity of the recall test in identifying factors affecting performance.

These findings do not support the results of studies done by Kendall, Mason, & Hunter (1980) and Rasool & Royer (1986) which found that there were lower proportions of correct responses on multiple-choice questions, cloze passages, recall and the Sentence Verification Technique tests when using expository passages than when they used narrative or fantasy passages. In this study, it was only during the recall test that the results aligned with the conclusions of previous studies.

One possible explanation for the discrepancies between prior research and the current study is the number of passages examined. In Kendall, Mason & Hunter (1980) and Rasool & Royer (1986), they used two sample texts of each story type for analysis, whereas in the current study each participant was tested with approximately 27 passages of each type. It is possible that with only two passages, the individual differences of each passage had a large effect on the results. Whereas with a larger sample size for text type, the individual differences of passages within a text type are averaged out and thus, reduce the chances of seeing a difference between the rates of response for the different text types. Also, the commonality of variance within text type categories indicates that there may be other variables affecting performance aside from general text type. There are other characteristics of the materials that may be having an effect on performance, such as, text topic, number of words or sentences and sentence structure. The existence of variance within text types suggests a possibility of finer categories of materials or material characteristics that effect performance.

Further, the measures used in the current study were rate of response, where prior research examined the proportion of correct responses. It is possible that one

measure is more sensitive to the effects of text type and could show differences that the other measure did not detect.

Tentatively, this study suggests that the type of text used to assess reading comprehension does not affect the rate of response, unless a recall test is being used as the assessment. Thus, this research suggests that assessments conducted by an instructor using one text type should be generalizable to other text types and the same should hold true for instruction. An instructor should be able to provide instruction using one text type and get improvements in performance in the other text type without direct intervention. However, since the results and conclusions of this study are contrary to previous research and common instructional practice, it's important to do further research.

This study also examined the effect of explicit time limits on various reading comprehension tests. Overall, the results show that for all three participants, imposing a time limit on the multiple-choice questions, cloze passages, sentence verification and recall tests does not affect the frequency distribution of their rate of correct and incorrect responses per minute (see Figures 5, 6 and 7). Similarly, none of the participants showed a consistent difference in celeration between the time limit and duration conditions for any of the reading comprehension tests (see Figures 8, 9 and 10). Interestingly, for the sentence verification test, Sally answered slightly more correct responses per minute in the time-limit condition than in the duration condition, but Harry and Lucy did not (See Figures 5, 6 and 7). For the recall test, Lucy recalled slightly more relevant words per minute in the time-limit condition than in the duration condition, but Harry and Sally did not (See Figures 5, 6 and 7). However, there was no difference

between the rate of "other" words recalled in the time-limit condition and that of the duration condition for all three participants.

Previous research (Lesaux, Pearson, & Siegel, 2006; Cates & Rhymer, 2006) has found that implementing and removing explicit time limits affects performance. In this study time limits do not affect the rate of correct or incorrect responses on any of the reading comprehension tests. However, there are many possible explanations for the discrepancies between Lesaux, Pearson, & Siegel (2006) results and the current study. First, the behavior under examination in the current study contained more component behaviors than the word-phrase reading examined by Cates & Rhymer (2006) or the performance on one standardized reading test as examined by Lesaux, Pearson, & Siegel, 2006. Since Lesaux, Pearson, & Siegel (2006) used only one standardized measure of reading comprehension, it contained only one test format as well, which in this case was a multiple-choice test. As discussed previously, different test formats require different behaviors; so testing across multiple tests formats would examine more behaviors than testing using just one test format.

Although Lesaux, Pearson, & Siegel (2006) also examined reading comprehension, they used raw number of correct responses as the gauge for effect. The time extension in the duration condition changes the number of opportunities to respond, thus making it difficult to isolate the explicit time limit as the only effect on performance. The current study used rate as the measure of comparison to eliminate the effect of extra opportunities to respond.

Another possible explanation for the discrepancy is the difference in instructions presented in prior research and in this experiment. In Cates & Rhymer (2006) they used

71

rate as the measure, but their intervention included telling the students to "go as fast as they can" only in the time-limit condition and presenting feedback of the student's rate of correct responses at the end of each timing. However, in the current study, there was no feedback of rate provided to the participants and the only instruction in both conditions was to try their best. The current study isolated the effect in the time-limit condition by maintaining all other aspects of the duration condition, including the instructions provided and the feedback delivered.

One consideration for these results is the amount of information received in the time limit and duration conditions. Since there was no difference in the frequency distributions of rates of response for the time-limit and duration conditions, more information is actually obtained through duration tests because the rate of response stays the same, but the participants can contact all of the stimuli on a test page rather than a small sample of questions/blanks. For example, since the rate of correct responses for the multiple choice test are the same in the time-limit and duration testing conditions, the time-limit test condition will provide a small sample of the participant's responses answering multiple choice questions, whereas in the duration testing condition, the participant would contact all of the questions on the test, providing more information.

Thus, one must consider the purpose of the assessments. Are the assessments being used to monitor progress or are they being used to gather information? If the assessments are being used to monitor progress, you want the most efficient measure of the behaviors the student has gained. Since there was no difference in rate between the time limit and duration tests, a time-limit test would show the same pattern of

behavior in the most efficient manner. As Bridgeman, McBride & Monaghan (2004)

point out in their pamphlet for the Educational Testing Service, a time limit can be an

important feature of an assessment when considering cost or instructional time. If the

test administrators are paid hourly, or if assessment time is limited, a time-limit test may

give you enough information for working purposes, but it will not provide as much

information overall as a duration assessment. Therefore, if the assessments were being

used to identify strengths and weaknesses, the duration test would provide more

opportunities and more information about component behaviors and deficits.

However, since our results did not align with those of previous research and

because time limits are used so frequently in test taking, more research is warranted to

isolate the effects of time limits on testing performance. Future research should explore

using rate to eliminate the effect of greater opportunities and take steps to isolate the

time limit as the only variable affecting responses.

To further examine the performance in the duration free recall tests, the

experimenter analyzed the participants' performance in the first 20 seconds of each

duration free recall. This allowed the experimenter to examine the duration recall

sessions for fatigue or change of rate when the recall was longer. By sampling the first

20-seconds of the long recall sessions and comparing it with the rate across the entire

recall session, a different picture of rate and celeration can be seen. The results

consistently showed a difference between the celerations seen in the first 20 seconds of

recall and the entire recall duration for 2 of the 3 participants (see Figure 11). The

celeration of Lucy's rates of relevant and "other" words recalled in both narrative and

expository conditions differed depending on whether they were derived from the first 20

seconds or the entire duration recall (see Figure 11). Also, in the expository condition, for Harry, although the rates of "other" words recalled remained similar in the first 20 seconds and the duration of the recall In the expository condition, his rate of "other" words recalled showed a steeper celeration in the first 20 seconds (x1.5) than they did during the measures taken in the entire recall timing (x1) (see Figure 11). However, Sally showed no differences in the rates or celerations of relevant and "other" words recalled in the first 20 seconds and those of the entire duration recall in both narrative and expository text conditions (see Figure 11).

Although no consistent pattern was seen, the presence of differences between the two measures for two of the participants further directs our attention to the size of the behavior sample, how we select the size of sample taken and how the mode and size of the sample can change the perceived skill level. When the experimenter changed the segment of analysis, a different pattern was identified and thus, would have yielded a different approach to intervention and placement.

Another goal of the study was to explore the relationship between practiced assessments of reading comprehension. In this realm, the findings are limited. The results of this study can only show that the highest rates of responses occurred during the recall test and the lowest rates of responses occurred during the cloze passages because the frequency distributions for recall were higher than those of the cloze passages, multiple-choice questions, and sentence verification sentence tests for all three participants (see Figure 5, 6 and 7). Also, the recall test was uniquely affected by text type for two out of the three participants.

The discrepancies between the effects of text type on the recall test, but not on the other tests for reading comprehension, support the idea that there are different component behaviors embedded in each reading comprehension test and thus each test can be effected uniquely. If there is not reciprocity between the different measures of reading comprehension, one explanation may be that different tests are actually detecting different, but related behaviors. In this instance, the recall may require a behavior that the other measures do not utilize.

Our results do not support the findings of Kendall, Mason and Hunter (1980) where the students showed higher proportions of correct responses in the multiple-choice questions than in the recall and cloze passages. It must be noted, however, that different measures were used during the current study than in prior research, so it is a bit like comparing apples to oranges. The current study examined rate, whereas the previous research looked at accuracy. Since the previous research did not measure rate of response it is difficult to compare the results of the current study with previous research.

However, this study and Kendall, Mason & Hunter's (1980) study highlight some significant complications in research comparing assessments of reading comprehension. One cannot address the relationship between the different tests for reading comprehension without acknowledging that each test requires a different type of response topography or channel of behavior and contains different stimuli that act on performance (Kendall, Mason & Hunter, 1980; Fletcher, 2006). Two complications arise from trying to compare tests with different forms of response: the channel of the response behavior and the response duration.

Each type of response in and of itself is a behavior. Thus, if you have a participant that has difficulty writing letters, the rate of correct responses for cloze passages will be affected, but not the participant's rate of words spoken during the recall test.

Additionally, employing different forms of response also means different response durations for each test. When measuring rate, the response duration has a large effect on the results that are obtained. For example, the correct response measured in the recall test was a relevant word spoken, whereas the correct response measured in the cloze passages was a corresponding word written. For all three of participants, it took a longer period of time to write a word than it took to speak a word. This difference in response duration needs to be taken into account when using rate to compare the reading comprehension tests. Otherwise, the low rate of responses in the cloze passages and the high rate of responses in the recall test could be mistaken for behavior deficits or task difficulty.

Also, each test under examination has different stimuli controlling the response behaviors. Any instruction presented prior to the test or the written aspects of the test itself can be stimuli that affect the participant's behavior. For the multiple-choice questions, each question presented is a stimulus that can affect the participant's responses. In fact, the instructor's behavior, be it the questions produced for the multiple-choice questions or the instructions delivered prior to the assessment, can exert some level of stimulus control over responding. Since there is no consensus as to what is important to ask or include in an assessment, each one ends up containing

different stimulus conditions. Thus, it is difficult to compare the information provided across reading comprehension tests or consider one assessment exhaustive.

As discussed previously, the majority of studies examining reading comprehension will use one isolated measure to assess an individual's skill level and deficits. The current study examined the effects of measuring with multiple assessments and whether multiple assessments provided more information than using one means to measure reading comprehension. The results indicate that more assessments did provide a broader range of information than using one assessment. However, without further analysis of the stimuli involved each reading comprehension test, our results can only conclude that the different tests measure differently, not what they measure.

Any conclusions that have been drawn from this comparison of reading comprehension tests should be incorporated with caution, and with consideration to the different response requirements of each tests, the measurement of the response, and the variety of behaviors required to perform each test. Tentatively, this study does show that measuring widely provided more information because the effects of text type and time limits were seen on some tests, but not others.

There were some interesting patterns brought out in this research. For one, all three participants showed some form of decline in rate of responding across the research study. Harry showed stable rates of response in the cloze passages, multiple-choice questions, and the sentence verification test, but had declining rates in the recall test (see Figure 5). On the other hand, Sally and Lucy demonstrated stable and declining rates in the cloze passages, multiple-choice questions, and SVT task, but had increasing rates in the recall test (see Figures 6 and 7). Also, anecdotally, all of the

participants began to comment to the experimenter that they wished the sessions were finished and that they were tired of doing all of the tests, especially when they were completing the cloze passages. One possible explanation for the declining performance could be a lack of a functional motivation system for the participants. The participants received a prize at the end of each session regardless of performance and received general praise throughout the session, so no contingency existed for correct responses or increases in rate during performance on the tests. No assessment was done to examine the effectiveness of selecting from the prize box or receiving general praise as reinforcers. Another explanation could be boredom with the procedures or materials. Although they were presented with different topics each session, there is no guarantee that boredom with the process or topics did not occur.

Another interesting consideration from this study is the participants selected for the study. Two of the three participants tested high on standardized intelligence assessments and were reported to perform well on class reading assignments. Both Harry and Sally showed very low rates of incorrect responses and high rates of correct responses across all of the reading comprehension tests. Since their errors are low, it is difficult to determine which, if any, of the reading comprehension tests more reliably identifies their deficits. As Cain & Oakhill state in their descriptive analysis of reading comprehension measures, "If two [individuals] obtain comparable scores but are performing at ceiling, it is always possible that they might obtain different scores if the task were more sensitive" (p. 704, 2006). Although Harry was performing at higher rates of correct responses, both participants were reaching a ceiling in term of correct responses and could have produced very different results if the stories and tests were at

a higher readability level. One consideration is that Harry, who performed a high rate of correct responses and a low rate of incorrect responses across all of the tests, may be a model of what we call having "good reading comprehension." It is possible that children who perform well across a wide range of measures are considered to have good comprehension. However, the only way to assess such things is to measure widely.

There were some limitations to the study. One limitation of the study was the number of variables being modified. In an attempt to examine multiple components of reading assessments, the experimenter manipulated several conditions at once (text type, test condition, and reading task). Since all of the variables were modified concurrently, it is difficult to tease out the effects produced by any one variable. The results were a conditional analysis for each variable, but could only tentatively identify a direct effect of any one variable. A simpler study, isolating one dimension of the assessment would yield more definitive results.

Another limitation of the study was the coding system used in the recall test. The coding system for "other" words recalled was very broad and included multiple types of errors (extra information not provided in the study, extraneous information about the testing sessions, and incorrect information). With so many types of errors coded as "other," it is difficult to identify the participants' deficits. All three participants show relatively high rates of errors, but anecdotally I can report that they were making very different kinds of errors. Often, Harry's errors were providing extra information about the reading topic. Whereas, Sally's errors were primarily filler words such as, "like" and "umm." Lucy's errors were primarily incorrect information. Without a better coding

system all of these errors appear the same and may inflate the error rate for the recall test.

Although a standard assessment does not currently exist, due to NCLB (2001), the pressure increases to find predictive, research-based assessments (for a definition of research based see NCLB, Section 9101.37), such that the teachers can preventatively gauge and instruct reading component behaviors prior to the students taking the state-wide tests (Bigger, 2006). Proper assessment and intervention is needed to positively affect reading comprehension improvements because the better the assessment, the more appropriate the intervention, and the greater the individual student's success. As previously discussed, there are different assessment methods, offering their own advantages and disadvantages. However, a test or assessment is only valuable if it measures what it claims to measure.

The results of this study encourage researchers to look more critically at how reading comprehension is measured and whether those measures are correlated with the individual target behaviors. Although, there is continued ambiguity as to an operational definition for reading comprehension, this study just cracks the surface of the research exploring the parameters of assessing the construct "reading comprehension." As seen in this experiment, different measures of reading comprehension appear to assess different component behaviors and are uniquely affected by text type and time limits on testing performance.

Perhaps reading comprehension is so difficult to define and consistently measure because it is actually a construct of many behaviors that when exhibited together are called "reading comprehension." Instead of searching for one consistent, sensitive

measure of reading comprehension, this research suggests that assessment should focus on measuring widely in an ability to address the many behaviors involved in obtaining information from text and aid in determining proper instruction.

The limitations this study encountered in trying to compare reading comprehension tests tells us that another avenue of approach to the exploration of the behaviors necessary to benefit from text could be less fraught with ambiguity. An approach focusing more on individual behaviors rather than a construct of a behavior could be a more informative methodology. Therefore, future research should continue to focus on the characteristics of assessment, but focus more on the individual skills and stimulus control involved in each type of reading comprehension assessment. It would benefit the research and practice to begin to discuss reading comprehension as a complex of behaviors, each one important and each one measured in its own way. Findings should be discussed in terms of specific behaviors instead of as a concept, such as, assessing the ability to recall a reading or answer questions about a text, as opposed to a behavior, reading comprehension. It is necessary to define and agree upon a taxonomy of behaviors, and the means by which to measure them, rather than simply letting our definitions remain artifacts of a muddled process, lacking consensus on the basic behaviors we are trying to ultimately teach.

Also, although the rate of correct and incorrect responses is an important level of information, to be able to formulate intervention or structure programming, a thorough error analysis would need to be completed. An error analysis is beyond the scope of this study, but it could be the next logical step. Finally, it is important for future research to examine the setting of time limits during assessments. In the current study, the time

limits were arbitrarily set based on pilot tests. Future research should focus on

determining time limits that provide enough information to design instruction, but keep

assessments efficient.

In conclusion, the results of this study remain tentative until research is done

examining the taxonomy of behaviors comprising "reading comprehension" and the

stimulus control involved in each test format. This study shows that a comparison

across different tests for reading comprehension will not bring us closer to a definition of

reading comprehension and a reliable measure, but it did prompt us to address the

differences seen between the assessments themselves and point us in a different

direction for future research.

APPENDIX A
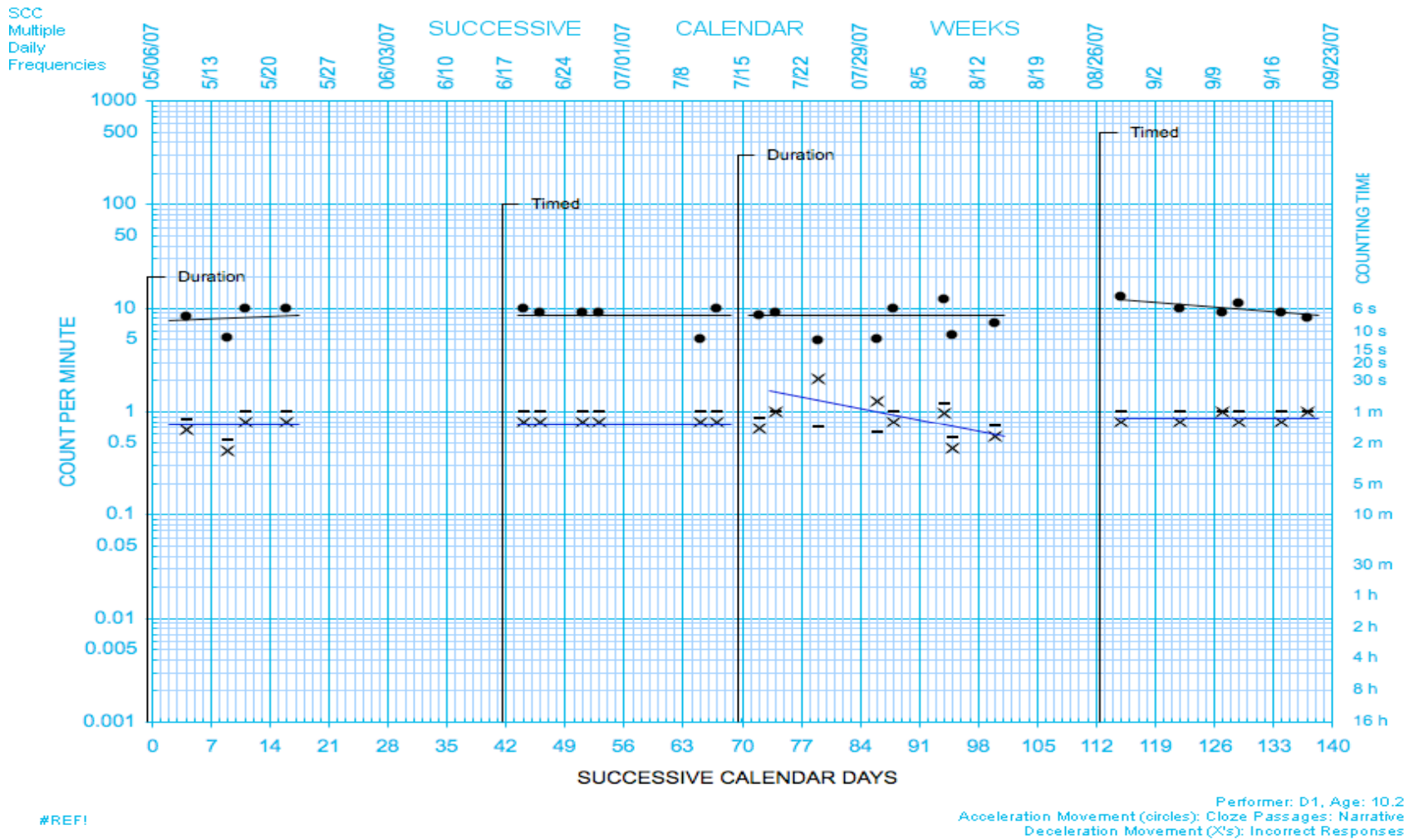
INDIVIDUAL DAILY GRAPH SAMPLES

Figure A.1. The rate of correct (solid dots) and incorrect ("X"s) responses per minute for Harry on the cloze passages in the time limit and duration conditions.
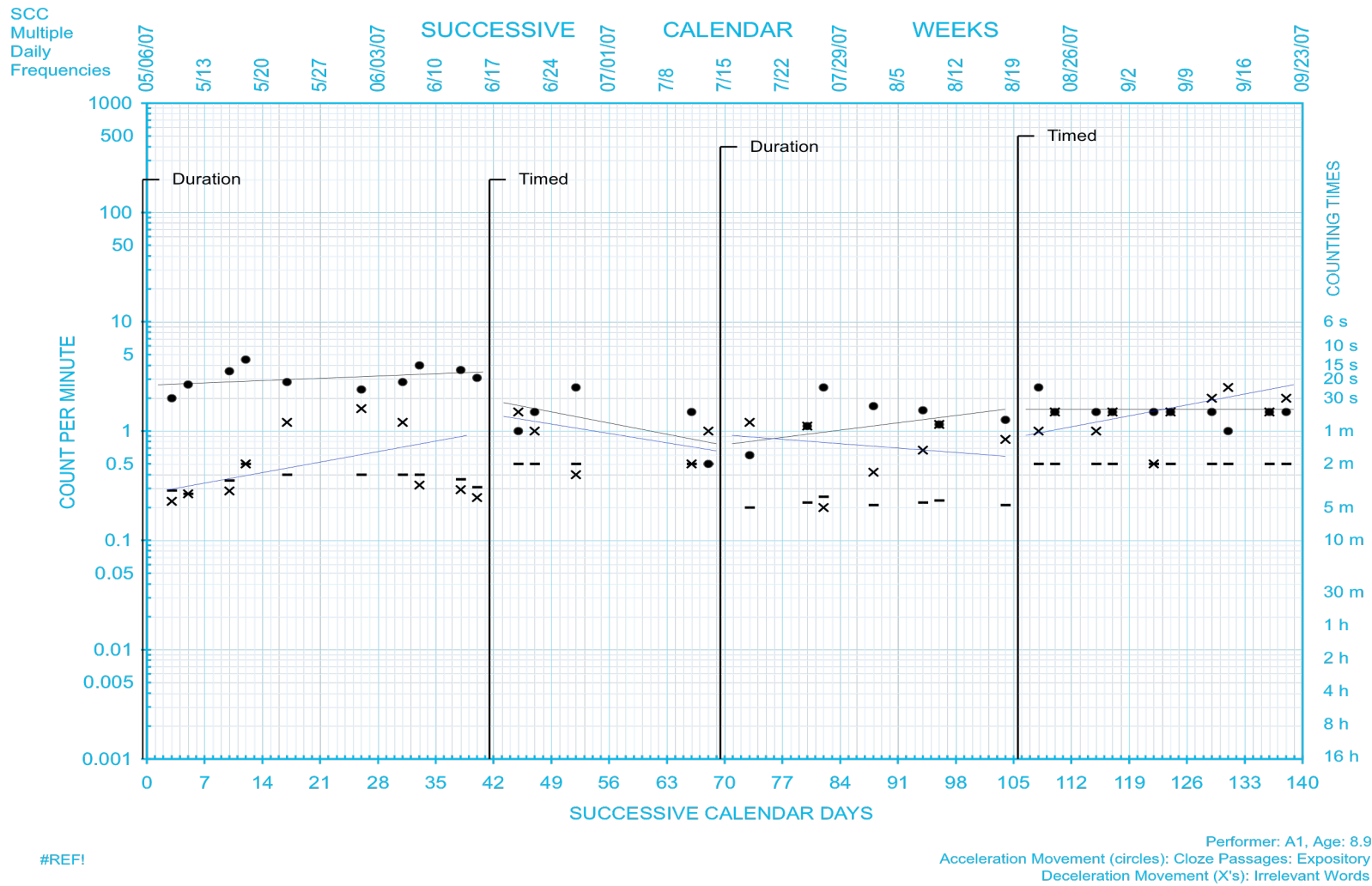
Figure A.2. The rate of correct (solid dots) and incorrect ("X"s) responses per minute for Lucy on the expository cloze passages in the time limit and duration conditions.

85

APPENDIX B

READING AND TESTING MATERIAL SAMPLES

Carnaval in Mazatlan

It was February and the week of winter vacation. Carmelita and her mother were flying to Mexico. They would stay with Carmelita's aunt and uncle, Tia Rosa and Tio Miguel. They planned the trip for the five days of Carnival in Mazatlan. The city hosted the biggest Carnaval celebration in all of Mexico.

"Mama, tell me again about Carnaval," begged Carmalita.

"You'll see it soon enough," Mama smiled. "People will wear masks and have all kinds of fun."

"Why do they wear masks?" Carmelita wondered.

"Mostly, they let people disguise themselves. That is a tradition of Carnaval."

Tia Rosa met them at the airport. "Welcome!" she exclaimed, hugging Mama and Carmalita in turn. She drove carefully from the airport to her home. They passed crowds of people dressed in costumes. Everyone in this town seemed to be thronging through the streets.

"I can't wait to join the party!" Carmelita yelled.

"Be sure to take some pictures with your new camera. Then you'll have something to show your classmates when we come home," Mama said.

The first few days were exciting. Tia Rosa and Tio Miguel took Mama and Carmelita to the outdoor festivals and dances. There was delicious food sold by street vendors. There was lively music to applaud.

But the weekend was the most wonderful part of the Carnaval celebration. Carmelita loved fireworks that sparkled over the waters. On Sunday, Mama and Tia Rosa taught Carmelita some new dance steps. They laughed like schoolgirls.

Then, Tio Miguel asked Carmelita to be his dance partner. They stepped and whirled through the crowds very fast. Carmelita could hardly catch her breath. "I wish every day could be just like carnival! She shouted above the noise of the music.

"Oh, once a year is quite enough for me," her uncle laughed. "Otherwise, my feet would become very sore." Tia Rose and Mama joined them and they all walked down by the water.

318 words

Adapted from *Targeting the TAKS: Reading, Writing, and Mathematics*
Copyright 2005 by Harcourt Achieve Inc. Used by permission.

Carnaval in Mazatlan
Circle the best possible answer to the question:

1. What was this story mainly about?
  (A) People wearing masks
  (B) a big celebration in Mexico
  (C) a family reunion
  (D) a little girl learning to dance
	2. Which of these is a fact in this story?
	  (A) Tio Miguel is old and tired
	  (B) Tia Rosa loves to dance
	  (C) Carmelita's favorite part is the fireworks
	  (D) Carnaval is celebrated in February

3. The word "thronging" appears in the story when the author says, "Everyone in this town seemed to be <u>thronging</u> through the streets." What do you think it means?
  (A) crowding
  (B) walking
  (C) looking
  (D) whispering
	4. Mama tells Carmelita to take some pictures so she can
	  (A) remember her trip
	  (B) bring them to school
	  (C) pin them on her wall
	  (D) use them for a class project

5. What is one difference between Carmelita and Tio Miguel?
  (A) Carmelita likes Carnaval, but Tio Miguel does not
  (B) Carmelita likes to dance, but Tio Miguel does not
  (C) Carmelita wants Carnaval to last forever, but Tio Miguel does not
  (D) Carmelita wants to leave Mexico, but Tio Miguel does not
 	
7. Which sentence from the story best shows how Carmelita feels about Carnaval?
  A. "Mama, tell me again about Carnaval…"
  B. The first few days were exciting.
  C. Carmelita saw people thronging the streets.
  D. "I wish every day could be just like Carnaval!"
 	
9. In the story, they call her "Tia Rosa." What do you think "Tia" means?
  A. Aunt
  B. Girl
  C. Uncle
  D. Boy
	6. What happened last in the story?
	  (A) people were thronging the streets
	  (B) they all went for a walk by the water
	  (C) Tia Rosa and Mama taught her how to dance
	  (D) Tia Rosa gave her a big hug
	
	8. What part of the Carnaval does Carmelita like best?
	  A. Flying on the plane because she likes to travel to new places.
	  B. The first day because of the dances, music and food.
	  C. The days before because she is excited about going on the trip.
	  D. The weekend because of the fireworks.
	
	10. How many days does Carnaval last?
	  A. three
	  B. five
	  C. ten
	  D. one

Carnaval in Mazatlan

airport	hosted
costumes	winter
sparkled	dances
Tio	sore
disguise	masks

Fill in the blanks with a word that completes the meaning of the sentence:

It was February and the week of _____ vacation. Carmelita and her mother were flying to Mexico. They would stay with Carmelita's aunt and uncle, Tia Rosa and _____ Miguel. They planned the trip for the five days of Carnaval in Mazatlan. The city _____ the biggest Carnaval celebration in all of Mexico.

"Mama, tell me again about Carnaval," begged Carmalita.

"You'll see it soon enough," Mama smiled. "People will wear _____ and have all kinds of fun."

"Why do they wear masks?" Carmelita wondered.

"Mostly, they let people _____ themselves. That is a tradition of Carnaval."

Tia Rosa met them at the _____. "Welcome!" she exclaimed, hugging Mama and Carmalita in turn. She drove carefully from the airport to her home. They passed crowds of people dressed in _____. Everyone in this town seemed to be thronging through the streets.

"I can't wait to join the party!" Carmelita yelled.

"Be sure to take some pictures with your new camera. Then you'll have something to show your classmates when we come home," Mama said.

The first few days were exciting. Tia Rosa and Tio Miguel took Mama and Carmelita to the outdoor festivals and _____. There was delicious food sold by street vendors. There was lively music to applaud.

But the weekend was the most wonderful part of the Carnaval celebration. Carmelita loved fireworks that _____ over the waters. On Sunday, Mama and Tia Rosa taught Carmelita some new dance steps. They laughed like school girls.

Then, Tio Miguel asked Carmelita to be his dance partner. They stepped and whirled through the crowds very fast. Carmelita could hardly catch her breath. "I wish every day could be just like carnival! She shouted above the noise of the music.

"Oh, once a year is quite enough for me," her uncle laughed. "Otherwise, my feet would become very _____." Tia Rose and Mama joined them and they all walked down by the water.

Carnaval in Mazatlan

Place an "O" next to the sentences that tell information from the reading
Place an "N" next to the sentences that tell information not in the reading or which tell different information than the reading.

Carnaval is a big celebration only held in Mazatlan.

Tio Miguel loves to dance, but his feet tire quickly because he is old.

Carmelita had so much fun at Carnaval that she wished it would happen every day.

Carnaval is a big celebration with food, dancing and fireworks.

Tia Rosa gave Mama and Carmelita a big hug when she picked them up at the bus station.

Tia Rosa and Mama taught Carmelita new dance steps.

They all took a walk down by the water.

Carmelita's favorite part of Carnaval was the weekend because of fireworks.

By the end of Carnaval, everyone was exhausted and ready to go home.

Mama told Carmelita to take lots of pictures, so that she could show them to her brother and father.

Carmelita and her mother were flying to Mexico.

This was Carmelita's first trip to Mexico and her first time to Carnaval.

They passed crowds of people dressed in costumes.

The Ideas of Ashoka

The Mauryan Empire of India existed over 2,200 years ago. It was a vast empire that covered thousands of miles. The empire had a very special ruler. His name was Ashoka.

In Ashoka's time, wars were common. Empires fought each other often. Ashoka led a fight against Kalinga. Kalinga was an area south of the Mauryan Empire. Ashoka won the battle and took over Kalinga. Most rulers would have been happy to win. But Ashoka was saddened by his victory. He saw how the people in Kalinga had suffered and he did not like it. He spoke out against the idea of using force to take over other countries. He promised never to do it again.

Ashoka spent his life teaching and helping people. He taught many people about the importance of honesty, compassion, and concern for others. This was not easy to do back then. There were no televisions, radios, or newspapers. In order to teach many people, Ashoka had his ideas carved on large rocks and pillars. These rocks and pillars were placed throughout his empire. When people traveled, they saw the rocks and pillars. They learned the teachings of their emperor. These monuments are called the Rock Edicts and Pillar Edicts. On them Ashoka asked his followers to lead a good life and help others.

Ashoka put many of his ideas about helping people into action. He created hospitals for both people and animals. He had trees planted along the busy roads to provide shade. He also had rest houses built for tired travelers. Ashoka asked everyone who worked for him to help the needy.

Ashoka died around 238 B.C. but the Rock Edicts and Pillar Edicts are still in India and other nearby countries. The messages on them to live a life of "little sin and many good deeds"- are as important today as they were thousands of years ago.

314 words

The Ideas of Ashoka

carved	won
fought	existed
good	shade
happy	ruler
vast	teaching

Fill in the blanks with a word that completes the meaning of the sentence:

The Mauryan Empire of India _____ over 2,200 years ago. It was a _____ empire that covered thousands of miles. The empire had a very special _____. His name was Ashoka.

In Ashoka's time, wars were common. Empires _____ each other often. Ashoka led a fight against Kalinga. Kalinga was an area south of the Mauryan Empire. Ashoka _____ the battle and took over Kalinga. Most rulers would have been _____ to win. But Ashoka was saddened by his victory. He saw how the people in Kalinga had suffered and he did not like it. He spoke out against the idea of using force to take over other countries. He promised never to do it again.

Ashoka spent his life _____ and helping people. He taught many people about the importance of honesty, compassion, and concern for others. This was not easy to do back then. There were no televisions, radios, or newspapers. In order to teach many people, Ashoka had his ideas _____ on large rocks and pillars. These rocks and pillars were placed throughout his empire. When people traveled, they saw the rocks and pillars. They learned the teachings of their emperor. These monuments are called the Rock Edicts and Pillar Edicts. On them Ashoka asked his followers to lead a _____ life and help others.

Ashoka put many of his ideas about helping people into action. He created hospitals for both people and animals. He had trees planted along the busy roads to provide _____. He also had rest houses built for tired travelers. Ashoka asked everyone who worked for him to help the needy.

Ashoka died around 238 B.C. but the Rock Edicts and Pillar Edicts are still in India and other nearby countries. The messages on them to live a life of "little sin and many good deeds"- are as important today as they were thousands of years ago.

The Ideas of Ashoka

**Place an "O" next to the sentences that tell information from the reading
Place an "N" next to the sentences that tell information not in the reading or
which tell different information than the reading.**

Ashoka was a very special servant.

Ashoka taught that you should live a good live and help other people.

Ashoka created hospitals for both people and animals.

Being an animal lover, Ashoka adopted several cats and dogs.

Travelers learned Ashoka's teachings from books left on the road.

Trees were planted along the road to provide shade.

Ashoka was a great gardener and planted flowers to adorn the roads.

The Rock Edicts and Pillar Edicts still exist in India to remind us of Ashoka's teachings.

Ashoka taught the important idea of helping other people and living a good life.

Ashoka's army lost the battle, but he learned an important lesson about how to threat others.

In Ashoka's time, wars were rare.

He taught many people about the importance of honesty, compassion, and concern for others.

The Ideas of Ashoka

Circle the best possible answer to the question:

1. The most likely reason that the author wrote this passage was to - Ⓐ tell about the Mauryan Empire Ⓑ explain why Ashoka was a special leader Ⓒ give information about Edicts Ⓓ persuade readers to be like Ashoka	2. Which of these is a fact in this story? Ⓐ Ashoka loved people and animals Ⓑ The roads got very hot for travelers Ⓒ Ashoka taught that it is important to help other people Ⓓ The rocks and pillars will never be destroyed
3. The story says, "It was a <u>vast</u> empire that covered thousands of miles." What does "<u>vast</u>" mean? Ⓐ wide-spread Ⓑ private Ⓒ small Ⓓ strong	4. What best describes Rock and Pillar Edicts? Ⓐ long-lasting and light Ⓑ heavy and long-lasting Ⓒ cheap and light Ⓓ heavy and cheap

5. Which of the following is the best summary of the passage?
 Ⓐ The Mauryan Empire of India existed over 2,200 years ago. It fell apart after Ashoka died.
 Ⓑ Ashoka was the emperor of the Mauryan Empire. He wrote Edicts on pillars and rocks. People still read them today.
 Ⓒ Ashoka was the emperor of the Mauryan Empire. He taught people to be kind. He had messages carved in rocks and pillars that taught people to be good
 Ⓓ Ashoka was a strong man who liked to fight other countries.
 	
 7. What is one of the ways that Ashoka is different from the other leaders of his time?
    A. He did not believe in using violence to control people
    B. He was able to read and write
    C. He refused to defend himself
    D. He was not a religious person
	6. Ashoka was not happy about taking over Kalinga because -
	 Ⓐ the people in Kalinga were not Buddhists
	 Ⓑ the war lasted a very long time
	 Ⓒ the country was smaller than he thought
	 Ⓓ the people in Kalinga had been hurt.
	
	 8. Which of the following is something that Ashoka would not likely do?
	    A. fund a school for poor children
	    B. help poor villagers get food
	    C. assist in the construction of a new public park
	    D. attack a government that did not like him

APPENDIX C

RECRUITMENT FLYER

Does your Child have difficulty with reading tests?

Do assignments involving reading create frustration and despair?

Want to know more about why they are so tough for your child?

Join a study being conducted by the Behavior Analysis Department at the University of North Texas

This study will look at your child's performance on different types of reading tests and examine possible avenues of instruction for your child

All meetings and sessions will take place at Harvest Christian Academy or a convenient location for the family

If you are interested in your child participating or would just like to find out more information, please contact the lead investigator.

Thank You!!

REFERENCES

Bigger, S. L. (2006). *Data-driven decision-making within a professional learning community: Assessing the predictive qualities of curriculum-based measurements to a high-stakes, state test of reading achievement at the elementary level.* Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, 127 pages.

Bonfiglio, C.M., Daly, E.J., Martens, B.K., Lin, L.H., & Corsaut, s. (2004). An experimental analysis of reading interventions: Generalization across instructional strategies, time, and passages. *Journal of Applied Behavior Analysis*, *37*(1), 111-114.

Bower, B. & Orgel, R. (1981). To err is divine. *Journal of Precision Teaching, 11*(1), 3-12.

Bridgeman, B., McBride, A. & Monaghan, W. (2004). *Testing and time limits. ETS R & D Connections,* Report from the Educational Testing Service, Princeton, NJ. Policy Information Center, 6 pp.

Brown-Chidsey, R., Davis, L. & Maya, C. (2003). Sources of variance in curriculum-based measures of silent reading. *Psychology in Schools, 40*(4), 363-377.

Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology, 21*, 335-351.

Cain, K & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*, *76*, 697-708.

Cain, K. Oakhill, J. & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31-42.

Cates, G.L. & Rhymer, K.N. (2006). Effects of explicit timings on elementary students' oral reading rates of word phrases. *Reading Improvement, 43*(3), 148-156.

Cutting, L.E. & Scarborough, H.S. (2006). Predictions of reading comprehension: Relative contribution of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, *10*, 277-299.

Daly, E., Bonfiglio, C., Mattson, T., Persampieri, M., & Foreman-Yates, K. (2005). Refining the experimental analysis of academic skills deficits: Part 1. An investigation of variables that affect generalized oral reading performance. *Journal of Applied Behavior Analysis, 38*(4), 485-497.

Fletcher, J.M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10*(3), 323-330.

Francis, D.J., Snow, C.E., August, D., Carlson, C.D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading, 10*(3), 301-322.

Frith, U., & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *British Journal of Developmental Psychology, 1,* 329-342.

Gustafsson, J.E. & Rosen, M. (2006). The dimensional structure of reading assessment tasks in the IEA Reading Literacy Study 1991 and the Progress in International Reading Literacy Study 2001. *Educational Research and Evaluation, 12*(5), 445-468.

Hall, W.S. (1989). Reading comprehension. *American Psychologist, 44*(2), 157-161.

Hosp, M.K. & Fuchs, L.S. (2005). Using CBM as an Indicator of decoding, word reading and comprehension: Do the relations change with grade? *School Psychology Review, 34*(1), 9-26.

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

Kendall, J. R., Mason, J.M. & Hunter, W. (1980). Which comprehension? Artifacts in the measurement of reading comprehension. *Journal of Educational Research, 73*(4), 233-236.

Kerstiens, G. (1986). *Time-critical reading comprehension tests and developmental students.* Paper presented at the Annual Meeting of the Educational Research Association*,* San Francisco, CA.

Kubina, R. M., & Starlin, C.M. (2003). Reading with precision. *European Journal of Behavior Analysis, 4*, 13-21.

Lahey, B. B., McNees, M.P., & Brown, C.C. (1973). Modifications of deficits in reading for comprehension. *Journal of Applied Behavior Analysis, 6*(3), 475-480.

Lesaux, N.K., Pearson, M.R. & Siegel, L.S. (2006). The effects of timed and untimed testing conditions on the reading comprehension performance of adults with reading disabilities. *Reading and Writing, 19*(1), 21-48.

Literacy statistics. Retrieved May 4, 2007, from http://www.readfaster.com/education_stats.asp

Lorch, E.P., O'Neil, K., Berthiaume, K.S., Milich, R., Eastham, D., & Brooks, T. (2004) Story comprehension and the impact of studying on recall in children with attention deficit hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology*, *33*(3), 506-515.

Lu, Y. & Sireci, S. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice*, *26*(4), 29-37.

Myles, B. S., Hilgenfeld, T.D., Barnhill, G.P., Griswold, D.E., Hagiwara, T., & Simpson, R.L. (2002). Analysis of reading skills in individuals with Asperger's syndrome. *Focus on Autism and Other Developmental Disabilities, 17*(1), 44-47.

Nation, K., Clarke, P., Marshall, C.M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research, 47*, 199-211.

Nation, K., & Norbury, C. (2005). Why reading comprehension fails: Insights from developmental disorders. *Topics in Language Disorders, 25*(1), 21-32.

Nation, K. & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*, 359-370.

Nation, K., & Snowling, M. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342-356.

Nesi, B., Levorato, M.C., Roch, M. & Cacciari, C. (2006). To break the embarrassment: Text comprehension skills and figurative competence in skilled and less-skilled text comprehenders. *European Psychologist, 11*(2), 128-136.

No Child Left Behind Act of 2001, Publ. No. 107-110, 115, Stat. 1425 (2002).

Norbury, C.F. & Bishop, D.V. (2002). Inferential processing and story recall in children with communication problems: A comparison of specific language impairment, pragmatic language impairment and high-functioning autism. *International Journal of Language Impairment & Communication Disorders, 37*(3), 227-251.

O'Conner, I. M. & Klein, P.D. (2004). Exploration of strategies for facilitating the reading comprehension of high-functioning students with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 34*(2), 115-127.

Ouellette, G.P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*(3), 554-566.

Pearson, P.D. & Hamm, D.N. (2005). The assessment of reading comprehension: A review of practices - Past, present, and future. In S.G. Paris & S.A. Stahl (eds.), *Children's reading comprehension and assessment* (pp. 13-69). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Poling, A., Methot, L.L., & LeSage, M.G. (1995). *Fundamentals of behavior analytic research* (pp. 72-77). New York: Plenum Press.

Rasool, J., & Royer, J. (1986). Assessment of reading comprehension using the sentence verification technique: Evidence from narrative and descriptive texts. *Journal of Educational Research, 79*(3), 180-185.

Rhymer, K.N., Evans-Hampton, T.N., McCurdy, M. & Watson, T.S. (2002). Effects of varying levels of treatment integrity on toddler aggressive behavior. *Special Services in School, 18*, 75-82.

Rhymer, K.N., Skinner, C.H., & Henington, C. (1998). Effects of explicit timing on mathematics problems: Completion rates in African-American third grade elementary students. *Journal of Applied Behavior Analysis*, *31*(4), 673-677.

Snyder, L., Caccamise, D., & Wise, B. (2005). The assessment of reading comprehension: Considerations and cautions. *Topics in Language Disorders, 25*(1), 33-50.

Storch, S.A. & Whitehurst, G.J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*(6), 934-947.

Vellutino, F.R., Tunmer, W.E., Jaccard, J.J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading, 11*(1), 3-32.

Wahlberg, T. (2001). Language development and text comprehension in individuals with autism. In *Autistic spectrum disorders: Educational and clinical interventions* (pp. 133-150): Elsevier Science Ltd.

Young, C. (2005). *The effects of timed readings on recall and comprehension in a child with Asperger's syndrome.* Unpublished master's thesis, University of North Texas, Denton.

U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), *Reading report card for the nation and the states*, 1992, 1994, 1998, 2002, 2003, 2005, and 2007 Reading assessments, retrieved November 8, 2007, from http://nces.ed.gov/nationsreportcard/nde/