

Abstract Title Page
Not included in page count.

Title: Using a New Reading Comprehension Assessment to Measure Discourse Representations and Identify Types of Comprehenders

Author(s): Sarah Carlson, Ben Seipel, & Kristen McMaster

Abstract Body

Limit 4 pages single spaced.

Background / Context:

Description of prior research and its intellectual context.

Many researchers focus on assessing the cognitive components of reading comprehension. However, researchers are challenged to find the best way to measure the cognitive components of reading comprehension because many reading comprehension assessments differ in terms of format (i.e., cloze, multiple-choice, open-ended); presentation (i.e., print); and type of cognitive components of reading the assessment is measuring (i.e., inferential processes) (Eason & Cutting, 2009). Additionally, most measures of reading comprehension are collected *after* reading (i.e., offline); yet, to comprehend a text, readers need to also build coherent mental representations *during* reading (i.e., online) (Graesser, Singer, & Trabasso, 1994). Finally, few traditional assessments are designed to diagnose specific comprehension difficulties or to identify types of comprehenders. Such information would be useful for teachers to plan instruction that addresses students' specific needs.

Assessing reading comprehension is vital to monitoring student growth, detecting comprehension difficulties, and testing theories of comprehension (Cain & Oakhill, 2006). Traditional assessments that assess comprehension (multiple-choice, true/false, open-ended, and modified cloze tests) have benefits and drawbacks. Multiple-choice tests are useful in group settings and are easy to administer; however, traditional multiple-choice tests require readers to choose only one correct choice and alternative choices are mainly distracters without theoretical meaning. True/false tests are also easy to administer, but every 'true' response can provide clues to the reader to generate a correct inference (Cain & Oakhill, 2006); it is impossible to indicate which type of inference the reader generates while reading. Open-ended questions can provide readers an opportunity to demonstrate that they made the correct inference; however, open-ended assessments are time consuming and difficult to score. Modified cloze tests, in which every 7th word is deleted and replaced with three options, have demonstrated to be a reliable form of assessment. However, they assess reading comprehension at a word integration level, not at a discourse level. Additionally, because cloze tests are timed, the task does not require a reader to build a complete mental representation of the text. Overall, these traditional assessments may not be ideal for teachers when making important instructional decisions.

Given these critiques of traditional reading comprehension assessments, the purpose of this study is to discuss findings for a new reading comprehension assessment. The development of this assessment was based on the benefits of Multiple-choice, Open-ended, and Cloze Comprehension Assessments (MOCCA). MOCCA measures reading comprehension at the discourse level, rather than at a word integration level, and may help identify readers with different reading comprehension abilities because each alternative choice has been developed with an inference type in mind. To explain further, online reading comprehension is assessed by having participants complete a cloze task with four choices at a discourse level. Instead of deleting every *n* number of words, one line of a short text is deleted. The deleted line occurs before the last line of the text such that readers have to infer and correctly choose an answer as they read the text to build a mental representation. One choice correctly completes the text coherently. The distracters can also complete the text, but do not complete the text coherently. In addition, the distracters were specially designed to identify different types of comprehenders, as well as identify subgroups of struggling comprehenders (McMaster et al., under review Rapp, et al., 2007).

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

The purpose of this study addresses the following questions: (1) What are the psychometric characteristics of this new reading comprehension assessment; (2) To what extent does this new reading comprehension assessment correlate with other reading measures; (3) To what extent does this new reading comprehension assessment distinguish among readers with different comprehension abilities; and (4) Does this new reading comprehension assessment distinguish between subgroups of struggling comprehenders?

Setting:

Description of the research location.

Students were from a large suburban school district in the Midwest, had parental consent, and assented to participation. Data collection took place in the students' classrooms for group testing and during individual sessions outside of the classroom in a designated area.

Population / Participants / Subjects:

Description of the participants in the study: who, how many, key features, or characteristics.

In this study, a total of 501 3rd-5th grade students participated in two study phases. Students were grouped into comprehension ability groups based on scores from three tests: (1) the Curriculum Based Measurement (CBM) Maze (Deno, 1985); (2) the district standardized reading comprehension assessment: The Computerized Achievement Levels Tests (CALT) (Northwest Evaluation Association, Lake Oswego, OR); and (3) Dynamic Indicators of Basic Early Literacy Skills, 6th Ed. (DIBELS) Oral Reading Fluency (ORF) test (Good & Kaminski, 2002). Scores from the CALT and DIBELS tests were provided by the school district. Students were then divided in 3 groups based on the percentile rankings for the CALT: Struggling comprehenders were in the 0-45% range ($n = 74$); average comprehenders were in the 55-70% range ($n = 70$); and good comprehenders were in the 80-95% range ($n = 74$). Students who did not fit these criteria were not selected for the individual testing portion of this study because their scores fit in between the above ranges and did not clearly fit in one group.

Intervention / Program / Practice:

Description of the intervention, program, or practice, including details of administration and duration.

For Track 2, this may include the development and validation of a measurement instrument.

Assessment development. The new reading comprehension assessment was designed with alternative answers that may provide insight to the different cognitive processes that take place during reading for readers with different reading comprehension skills, including the two types of struggling comprehenders mentioned above. There are no existing instruments designed that are suited to distinguish between comprehension differences such as the differences mentioned above. This assessment consists of questions based on principles behind multiple-choice, open-ended, and cloze comprehension questions. Thus, the name of the new assessment is **Multiple-choice Open-ended Cloze Comprehension Assessment (MOCCA)**. A sample text with choices from the MOCCA is provided in Appendix B.

The MOCCA is a paper-and-pencil cloze task with four answer choices at a discourse level. Instead of deleting every n number of words, one line of a short text is deleted. The deleted line always occurs before the last line of the text. Readers are required to choose one of four choices to complete the deleted line from the text. Choices were developed with different comprehender types in mind. The types of responses developed for the MOCCA are (1) Correct

response: completes the story with a causal and cohesive statement to build a coherent representation of the story; (2) Summary response: completes the story by paraphrasing the previous sentences presented in the story; (3) Local response: completes the story only at a local level, that is, the response is consistent and matches information from the immediately preceding sentence in the story; and (4) Distracter response: does not complete the story and is not inconsistent with developing a coherent representation of the story, however, elaborates the story. The average Flesch-Kincaid grade level for the texts is 4.4. Each text has a title, is seven sentences long, and is an average of 80.5 words.

Individual testing materials. Students participated in three additional reading activities that measured different components of reading (i.e., decoding, inference generation, and working memory). To measure decoding, students were administered the Woodcock Johnson III Word Identification and the Word Attack subtests (WJ ID and WJ Attack) (Woodcock et al., 2001). To measure inference generation students were administered a think-aloud task, in which readers are asked to read a text out loud, one sentence at a time, and then asked to say whatever came to mind when reading the sentence. This task has been used by researchers as a means to examine cognitive processes that occur during reading (e.g., inference generation) (e.g., Ericsson & Simon, 1993). Students read two texts for the think-aloud task. Each text was 7-15 sentences long, with a Flesch-Kincaid reading level of 4th grade. Finally, to measure working memory, students were administered a reading span task. This reading span task has been modified to be administered to children (modified by Swanson, et al., 1989 from a task developed by Daneman & Carpenter, 1980) and is correlated with other working memory tasks, convergent validity ($r=.70, p<.0001$). This task measures both the storage and processing components of WM. The task materials consist of a total of 28 test sentences and 6 practice sentences. There are two to five sentences in each set and each sentence ranges between six and ten words long. All of the final words to be repeated are nouns and none of the words are repeated in any of the sets. Last, a comprehension question is asked for one of the sentences within each set of sentences and each required answer is a noun.

Research Design:

Description of the research design.

The research design for this study is quasi-experimental.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.

For Track 2, this may include the use of existing datasets.

Group testing. For this study, there were two testing phases. In the first testing phase of the study, students were asked to complete two group-administered reading comprehension tests: (1) CBM Maze (10 minutes); and (2) MOCCA (20 minutes). During the CBM Maze task, students were given one minute read 3 texts. Whenever students came to a section in the text where a word was deleted and three options of words were inserted, students were asked to circle the word that belonged in the sentence. During the MOCCA test, students were asked to read short 7-sentence texts with one missing sentence. Each text had four multiple-choice responses to complete the missing sentence. Students were asked to read each text and decide with choice below each text best completed the story. Students were given 20 minutes to complete up to 40 items.

Individual testing. For the second testing phase of the study, students attended a one-on-one session with the researcher in which they completed the WJ Word ID and the Word Attack

subtests; the think-aloud task; and the working memory reading span task. All the individual sessions lasted approximately 30 minutes. Individual testing was tape recorded for future data coding and scoring. Accommodations were made for people with special needs such as physical disability or English as a second language.

Findings / Results:

Description of the main findings with specific details.

Question 1. To find the psychometric characteristics of the MOCCA. Coefficient alpha reliabilities (Cronbach's alpha) indicate that the MOCCA test yielded reliable scores for the correct and summary responses: MOCCA Correct Responses $\alpha = .927$; MOCCA Summary Responses $\alpha = .846$; and marginally reliable scores for the local and distracter responses: MOCCA Local Responses $\alpha = .645$; MOCCA Distracter Responses $\alpha = .594$.

Question 2. To determine whether the MOCCA correlates with other reading assessments. Correlation coefficients indicate that the MOCCA test is significantly correlated with several other reading measures: CALT $r = .580, p < .001$; CBM Maze $r = .325, p < .001$; DIBELS $r = .289, p < .001$; Woodcock Johnson III ID $r = .370, p < .001$; and Woodcock Johnson III Attack $r = .252, p < .05$). A complete correlation matrix is provided in Appendix B, Table 1.

Question 3. To determine whether the MOCCA can distinguish between readers with different comprehension abilities. Results from a MANOVA indicate a statistically significant main effect of comprehender type and question response type: $F(3, 98) = 7.954, p < .001, \eta^2 = .306$. Good comprehenders chose the correct answer more than both the average and poor comprehenders. Poor comprehenders chose the distracter, local and summary answers more than the good comprehenders, but only chose the distracter and local answers more than the average comprehenders. A graph depicting these results is provided in Appendix B.

Question 4. To determine whether the MOCCA can distinguish between subgroups of struggling comprehenders. First, K-means cluster analysis was conducted to determine whether struggling comprehenders form or cluster into two groups using the think-aloud data, as seen in previous research (see McMaster et al., under review). Results from the cluster analysis indicate statistically significant cluster groups of struggling comprehenders: (1) paraphrasers ($F = 26.33, p < .001$); and (2) invalid elaborators ($F = 42.32, p < .001$), which replicated previous findings. K-means cluster analysis was also conducted to determine whether struggling comprehenders form or cluster into two groups using the MOCCA data. Results from the cluster analysis indicate moderately significant cluster groups of struggling comprehenders (1) paraphrasers ($F = 4.69, p = .038$); and (2) distracters ($F = 3.30, p = .079$). We further ran a chi-square analysis using results from both cluster analyses. Chi-square results did not yield statistically significant differences between the think-aloud and MOCCA data. However, our sample sizes were extremely low for the struggling comprehender group and may have been a cause of not yielding statistically significant differences.

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

The results of our study show that findings from the MOCCA produce reliable and criterion-valid scores. In addition, we feel this test provides support for identifying different types of cognitive components comprehenders use during reading, as measured by the types of choices provided to complete each text. Teachers can use such information to plan instruction that addresses students' specific needs.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

- Cain, K. & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*, 76, 697-708.
- Daneman, M., & Carpenter, P. A. (1980). Individual difference in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Deno, S. L. (1985). Curriculum-Based Measurement: The emerging alternative. *Exceptional Children*, 52, 219-232.
- Eason, S. H. & Cutting, L. E. (2009). Examining sources of poor comprehension in older poor readers: Preliminary findings, issues, and challenges. In R. K. Wagner, C. S. Schatschneider, & C. Phythian-Sence (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension* (pp.263-283). New York, NY: Guilford.
- Ericsson, K. A., & Simon, H. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills (6th Ed.)* Eugene, OR: Institute for the Development of Educational Achievement.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-395.
- McMaster, K., van den Broek, P., White, M.J., Rapp, D.N., Kendeou, P., Espin, C., Carlson, S., & Bohn-Gettler, C. (under review). Identifying Interventions for Students who Struggle with Reading Comprehension Based on Online and Offline Reading Profiles.
- Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A., (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, 11, 289-312.
- Swanson, H. L., Cochran, K. F., & Ewers, C. A. (1989). Working memory in skilled and less skilled readers. *Journal of Abnormal Child Psychology*, 17, 145-156.
- Woodcock, R., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities and Achievement*. Itasca, IL: Riverside Publishing.

Appendix B. Tables and Figures

Not included in page count.

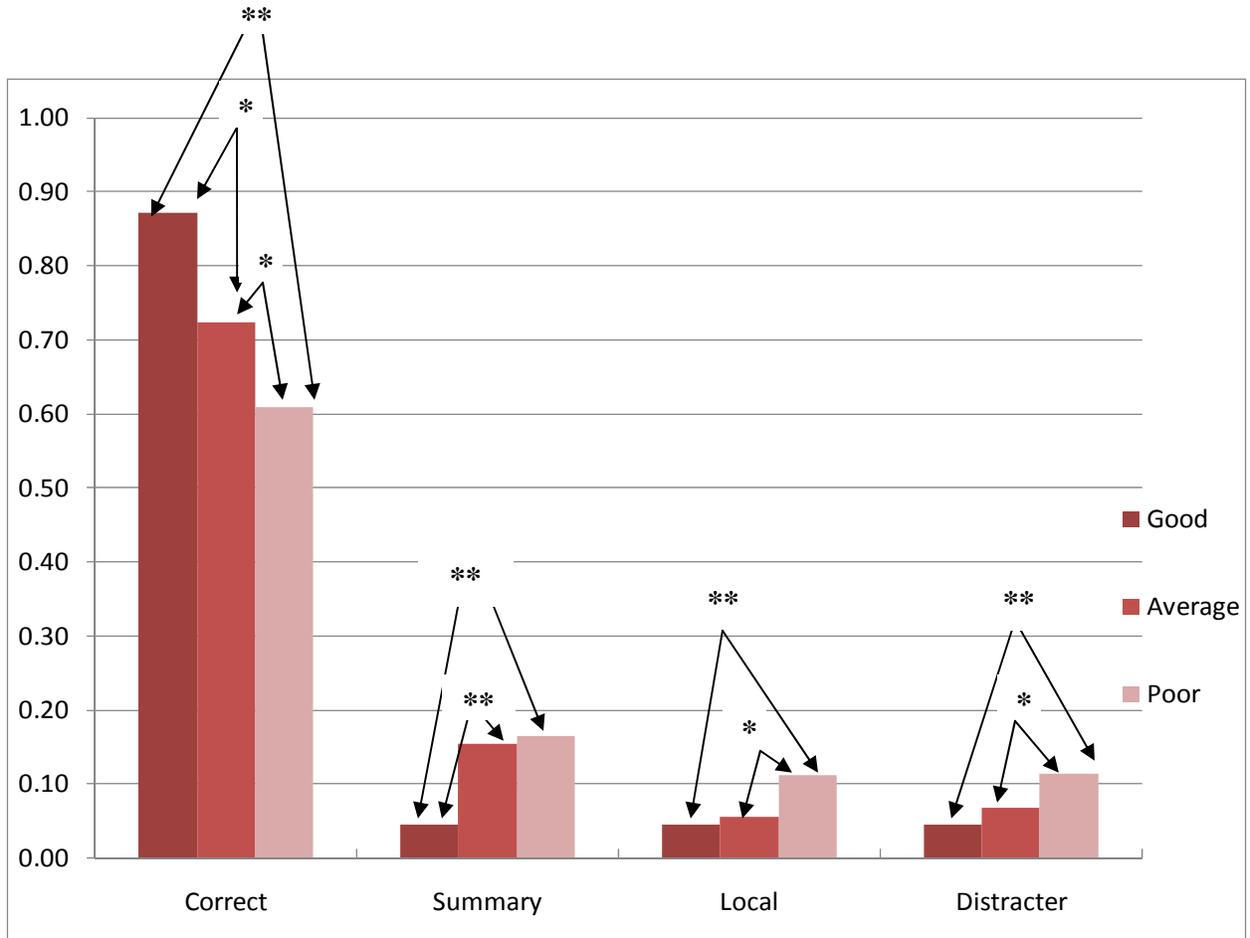
Table 1

	CALT Total %	MAZE Total	DIBEL Winter	Word ID	Word Attack	WM Words	MOCCA Total	MOCCA Correct	MOCCA Distract	MOCCA Local	MOCCA Summary
CALT Total %	1										
MAZE Total	.587**	1									
DIBEL Winter	.363**	.573**	1								
Word ID	.475**	.579**	.555**	1							
Word Attack	.323**	.398**	.367**	.752**	1						
WM Words	.271**	.369**	.177	.188	.195	1					
MOCCA Total	.318**	.554**	.413**	.301**	.214*	.259*	1				
MOCCA Correct	.580**	.325**	.289**	.370**	.252*	.144	-.133	1			
MOCCA Distract	-.452**	-.248*	-.233*	-.265**	-.195	-.026	.177	-.682**	1		
MOCCA Local	-.454**	-.263**	-.153	-.286**	-.223*	-.150	.139	-.700**	.499**	1	
MOCCA Summary	-.405**	-.206*	-.216	-.273**	-.171	-.121	.071	-.826**	.270	.288	1

* $p < .05$

** $p < .001$

Main Effects of MOCCA Response Type by Comprehender Type



* $p < .05$

** $p < .001$