

MOCCA College: An Assessment of Inferential Narrative and Expository Comprehension

Mark L. Davison¹, Ben Seipel², Virginia Clinton³, Sarah E. Carlson⁴, Patrick C. Kennedy⁵

¹Department of Educational Psychology, University of Minnesota, USA; ²School of Education, California State University, Chico, USA, ³Department of Education, Health, and Behavior Studies, University of North Dakota, USA, ⁴Department of Educational Psychology, Georgia State University, USA, ⁵Center on Teaching and Learning, University of Oregon, USA.

Abstract

MOCCA-C is an assessment of adult reading ability designed for early diagnosis of reading problems, for formative assessment in reading intervention planning, for assessment of reading improvement over time, and for assessment of reading intervention outcomes. It uses both narrative and expository reading passages and it currently has four forms. Two goals of this research were to compare narrative and expository passages on (a) their difficulty and (b) their ability to discriminate between good and poor readers. An additional goal was to assess whether narrative and expository passages measure the same or different comprehension dimensions. A final goal was to assess the reliability of forms. We randomly assigned students to forms with between 274 – 279 college students per form. Across the several forms, results suggest that narrative passages are easier and better discriminate between good and poor readers. However, both narrative and expository passages measure a single dimension of ability. MOCCA-C scores are reliable. Implications for research and practice are discussed.

Keywords: *Reading comprehension; reading assessment; adults; college students; formative assessment; diagnosis of reading problems.*

1. Introduction

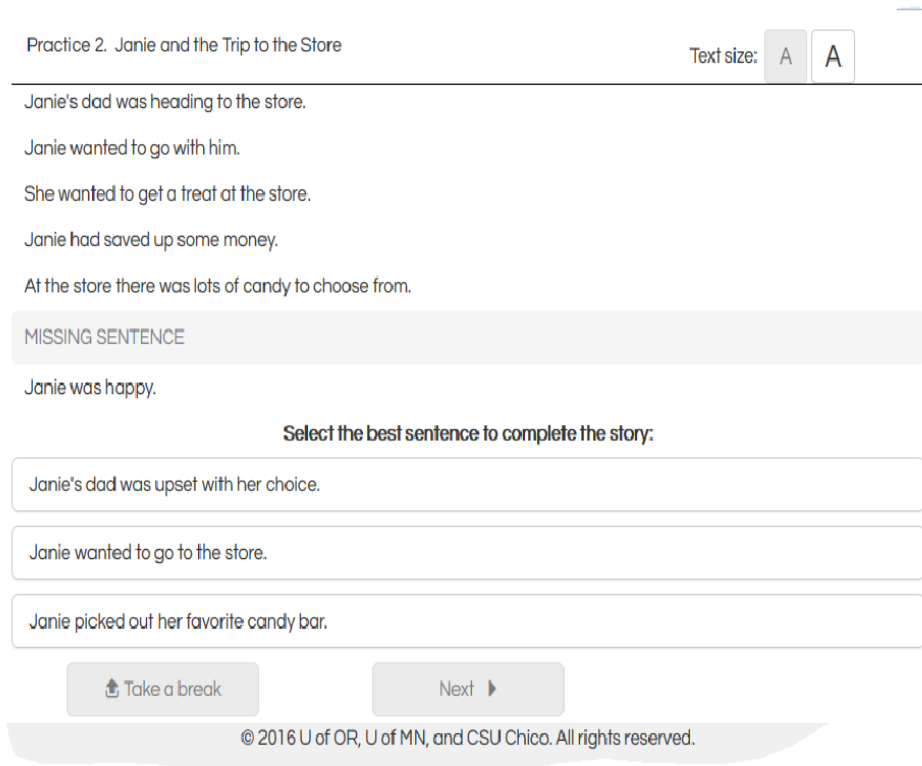
In the U.S., there has been increasing concern about the reading readiness of college students. The concern stems, in part, from the low percentage students meeting the ACT benchmark for college readiness (ACT, 2014). In 2013-2014, only 44% of high school graduates who took the ACT met the ACT benchmark for reading readiness (ACT, 2014). Moreover, approximately half of community college students could be considered struggling comprehenders—they have basic reading skills, but have difficulty generating appropriate inferences (Hoachlander et al., 2003). This has led us to pursue development of an inferential reading test for college students (a) to identify students in need of a reading intervention, (b) as a formative assessment for planning such an intervention, (c) to measure improvement during an intervention longitudinally over time, and (d) as an outcome measure.

MOCCA-C is based on earlier work to develop a reading assessment for students in grades 3 – 5 (Biancarosa et al., 2019; Davison et al., 2018; Liu et al., 2019). Unlike the earlier test that contained only narrative passages, the adult MOCCA contains both expository and narrative passages to reflect the expository nature of most college texts. It has multiple forms and therefore could be administered multiple times during an intervention to monitor student progress without the student having to take the same form twice. By administering forms before and during an intervention, the instructor may be better able to plan and adjust instruction as the intervention proceeds.

MOCCA-C is designed to be diagnostic of student errors. Each item consists of a paragraph with a sentence missing. From three alternatives, the student must select the sentence that best completes the story when inserted for the missing sentence. Figure 1 shows a sample item. Whereas most multiple-choice tests have two types of responses, each MOCCA-C has three types of responses, one correct response and two types of incorrect responses. The correct response is the causal coherent (CCI) response. The causal coherent response involves an inference that best completes the story line when inserted as the missing sentence.

The incorrect responses are drawn from observations of common error types in think-aloud research (e.g., Coté, Goldman, & Saul, 1998; McMaster et al., 2012.). The first type of incorrect response is a *paraphrase* (PAR), a sentence that simply repeats prior information from the text. Paraphrases do not involve an inference, do not move the story along by adding new information, nor do they complete the story line (narrative) or line of thought (expository). The second type of incorrect response is an *elaboration* (ELA). An elaboration involves an elaboration of, association with, or evaluation of information in the story. It can involve an inference and it goes beyond the explicit information in the story, but it does not complete the story line (narratives) or line of thought (expository). The answer types lead to three scores: a number correct score, a number of paraphrase response score, and a number

of elaboration response score. Since there are 50 items in each form, these three scores add to 50 if the student has answered every item.



Practice 2. Janie and the Trip to the Store

Text size: A A

Janie's dad was heading to the store.

Janie wanted to go with him.

She wanted to get a treat at the store.

Janie had saved up some money.

At the store there was lots of candy to choose from.

MISSING SENTENCE

Janie was happy.

Select the best sentence to complete the story:

Janie's dad was upset with her choice.

Janie wanted to go to the store.

Janie picked out her favorite candy bar.

Take a break

Next ▶

© 2016 U of OR, U of MN, and CSU Chico. All rights reserved.

Figure 1. Sample item.

MOCCA also yields a comprehension rate score, minutes per correct response. According to automaticity theory (LaBerge & Samuels, 1974) as comprehension improves, the comprehension becomes more automatic and faster. Automaticity may improve learning from reading material, because once comprehension becomes automatic, the reading process demands little conscious attention and does not interfere with a focus on the content to be learned from reading. There is a fifth score, number of items not reached that can be inferred from the CCI, PAR, and ELA scores given that the test has 50 items. The goal of this research was to examine the reliability, difficulty, and discrimination of the items.

2. Methods

The sample, test, and administration procedures are described only briefly here.

2.1. Sample

Since there are four forms there were four samples composed of 274, 279, 279, and 278 college students. The students constituted convenience samples from several states and several higher education institutions.

2.2. Instrument

Each form of the test contained 50 items with approximately equal numbers of expository and narrative items. Forms were matched on factors such as average number of sentences per item, sentence length, and Flesch-Kincaid readability.

2.3. Procedures

Participants were recruited through emails, social media, and courses in which instructors shared recruitment information. They participated for course credits or gift cards. Participants were randomly assigned to one of the four forms. All students took the test on a laptop or tablet. The computer administration included extensive instructions and showed two sample items. Students can go to the next item only after having answered the current item. If a student answers in less than 10 seconds, the answer is not accepted and they are told to read the item carefully before answering. There was no time limit on the test, although when the test was given in a class setting, the length of the class period may have set a limit. In other class settings, the instructor may have set a limit.

3. Results

Results are divided into four sections: descriptive statistics, reliability, difficulties and discriminations of narrative and expository items, and dimensionality of narrative and expository items.

3.1. Descriptive Statistics

Table 1 shows the descriptive statistics for the number correct (CCI), number of paraphrase (PAR), number of elaboration (ELA), and not reached (NR) items by form. While results varied by form, students generally answered about 80% of items correctly. When students failed to get credit for an item, it was somewhat more often because they did not reach the item. These trends are consistent across forms.

Table 1. Descriptive statistics for correct, paraphrase, elaboration scores, and not reached scores.

	Correct		Paraphrase		Elaboration		Not Reached	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Form 1	40.85	9.37	4.94	5.50	5.01	4.24	6.42	12.06
Form 2	39.14	8.73	6.06	5.50	4.80	4.10	6.76	12.74
Form 3	39.77	10.44	5.41	5.49	4.87	5.47	6.00	12.21
Form 4	41.64	8.28	3.90	4.83	4.46	3.97	6.74	12.63

3.2. Reliability

Table 2 shows the reliability for each of the scores. The reliability of the number correct scores are excellent, all above .90. Those for the Paraphrase and Elaboration scores are good to excellent, all but one above .80. The reliabilities for the Not Reached responses are high, but undoubtedly inflated by the non-independence between not-reached items at the end of a test.

Table 2. Reliability.

	Correct	Paraphrase	Elaboration	Not Reached
Form 1	.936	.875	.807	.981
Form 2	.917	.855	.771	.984
Form 3	.948	.863	.876	.974
Form 4	.932	.870	.809	.983

3.3. Difficulty and Discrimination of Narrative and Expository Items

Figure 2 shows the mean item difficulty (proportion correct) by form for narrative and expository items. For every form, the average item proportion correct is higher for the narrative items than for the expository items. To test this difference for significance, we performed a two-way ANOVA with item as the unit of analysis, with the factors of form and narrative vs. expository, and with item proportion correct as the dependent variable. The test statistic ($F_{1,192} = 266.165, p = .001$) would lead to rejection of the null hypothesis that the average item difficulty was equal for both narrative and expository items. We employed a

Type III sums of squares, thereby controlling for both the Form and Form x Narrative interaction in the hypothesis test.

Figure 3 shows the mean item-total correlation (a standard measure of item discrimination) for narrative and expository items by form. The average discrimination index is higher for the Narrative items across all forms. Again we performed a two-way ANOVA (Form by Narrative vs. Expository) with item as the unit of analysis and item discrimination as the dependent variable to test the hypothesis that the average item discrimination is equal for

narrative and expository items. The obtained $F_{(1,192)} = 19.781$, $p = .021$ would lead to rejection of the overall null hypothesis. The error bars in Figure 3 suggest that the difference is significant for all but Form 2.

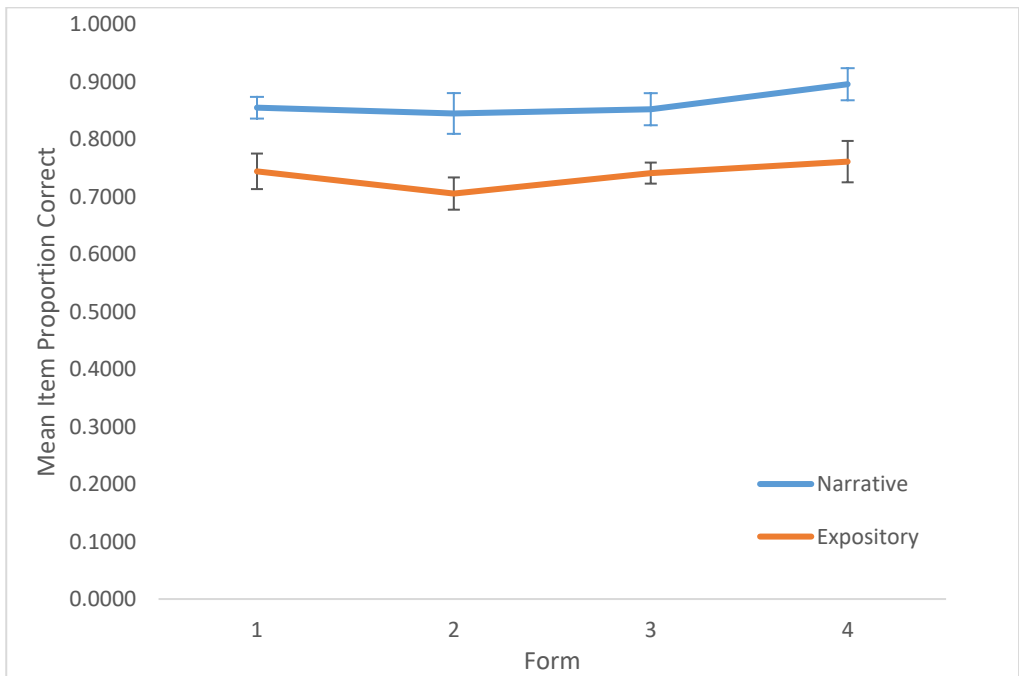


Figure 2. Mean proportion correct for narrative and expository items by form with 95% confidence intervals for the means.

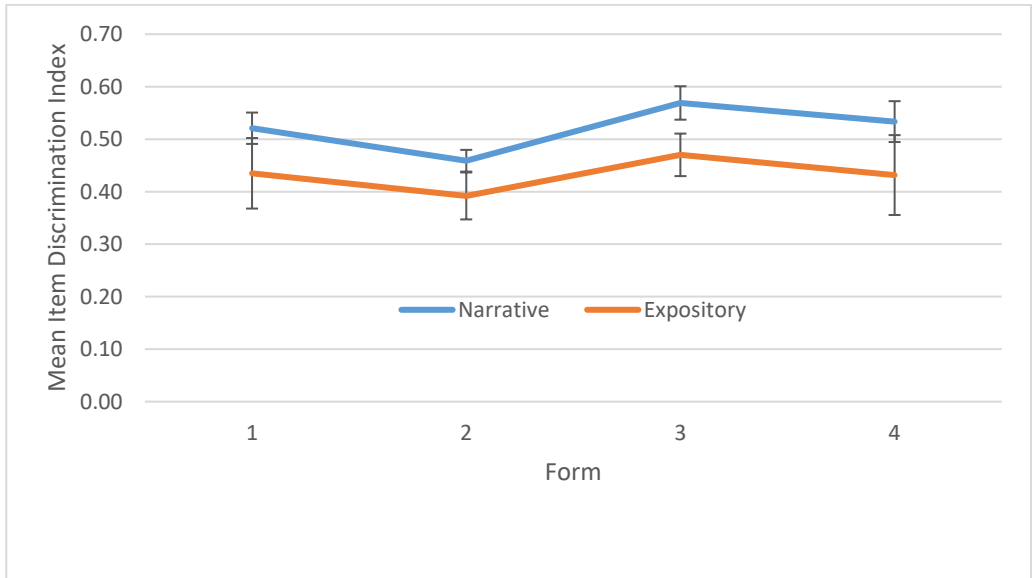


Figure 3. Mean item discrimination for narrative and expository items by form with 95% confidence intervals.

3.4. Dimensionality

Lastly, we used item response theory to address the question of whether the reading comprehension dimension underlying the narrative responses was the same as the dimension underlying the expository responses. To do so, we first fit a unidimensional, three-parameter logistic (3PL) model with all guessing parameters constrained equal for all 50 items. Then we fit a two-dimensional 3PL model with all guessing parameters constrained equal with narrative items discriminating only on the first dimension and expository items discriminating only on the second dimension. Table 3 show the statistics used to compare the models.

Table 3. Statistics comparing one- and two-dimensional models for the narrative and expository items.

	Correlation	AIC 1 Dimension	AIC 2 Dimension	BIC 1 Dimension	BIC 2 Dimension	Likelihood Ratio Test
Form 1	1.00	8757.64	8758.90	9122.57	9127.44	0.74
Form 2	1.00	9857.84	9859.87	10224.59	10230.25	0.01
Form 3	0.97	9236.64	9234.09	9603.39	9604.48	4.55*
Form 3	0.99	7536.29	7537.25	7902.67	7907.27	1.04

*p < .05

The IRT estimates of the correlations between the Narrative Dimension 1 and the Expository Dimension 2 are all at or above .97, suggesting that the two dimensions are virtually identical. The likelihood ratio statistic (LRT) provides a test of the null hypothesis that the two models fit equally well. It is not significant ($p > .05$) for all but Form 3. We can only reject the null hypothesis of equal fit for one of the forms, Form 3. The AIC is better (lower) for the unidimensional model for all but Form 3. The BIC is better (lower) for the unidimensional model on every form. With the exception of the Form 3 AIC and likelihood ratio test, results suggest that a single dimension underlies both the narrative and expository responses.

4. Discussion and Conclusions

Results lead to four major conclusions. Scores on the test have high reliability. The narrative items are easier, and they are somewhat more discriminating than are the expository items. Even though most college reading assignments involve expository text, narrative passages are just as useful as expository passages in measuring the comprehension ability required of college students.

In prior research (Graesser, McNamara, Cai, Conley, Li, & Pennebaker, 2014), authors have also found that expository text tends to be more difficult to comprehend. In part, this is because expository text contains technical vocabulary and relies more heavily on prior knowledge. In MOCCA-C, however, we have avoided technical vocabulary and the need for prior knowledge. Therefore, technical language and prior knowledge cannot explain the greater difficulty of expository items. Based on our experience writing items, it is our conjecture that the causal structure in expository text tends to be more subtle than that in most narrative passages, thereby making the expository texts more difficult.

Research on individualizing reading instruction based on MOCCA-C is at an early stage. McMaster et al. (2012) and Rapp et al. (2007) conclude that those who predominantly paraphrase and those who predominantly elaborate may benefit from different questioning strategies. In these studies, paraphrasers benefitted more from a questioning strategy emphasizing general connection making (e.g., “Make a connection to what you previously read.”), whereas elaborators benefitted more from a questioning strategy more narrowly focused on causal connections (e.g., “Why was Janie happy?”). However, a later study (McMaster, Espin, & van den Broek, 2014) using small group instruction did not replicate these earlier results, perhaps because small group instruction provides more optimal, individualized feedback about students’ comprehension or lack of comprehension.

References

- ACT, Inc. (2014). *The condition of college and career readiness*. Iowa City, Ia: Author.
- Biancarosa, G., Kennedy, P. C., Carlson, S. E., Yoon, H.-J., Seipel, B., Liu, B., & Davison, M. L. (2019). Constructing subscores that add validity: A case study identifying students at risk. *Educational and Psychological Measurement, 79*(1), 65 - 84. DOI 10.1177/0013164418763255
- Davison, M. L., Biancarosa, G., Carlson, S. E., & Liu, B. (2018). Preliminary Findings on the Computer Administered Multiple-choice Online Causal Comprehension Assessment (MOCCA), a Diagnostic Reading Comprehension Test. *Assessment for Effective Intervention, 2018, 43*(3), 169 – 181. DOI 10.1177/1534508417728685
- Cote, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processing, 25*(1), 1 – 53.
- Graesser, A., McNamara, D., Cai, Z., Conley, M., Li, H. and Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal, 115*(2), pp.210-229.
- Hoachlander, G., Sikora, A. C., Horn, L. & Carroll, C. D. (2003). *Community college students: Goals, academic preparation, and outcomes*. Washington DC: National Center for Education Statistics, Institute of Education Sciences, U. S. Department of Education.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293–323. doi:10.1016/0010-0285(74)90015-2
- Liu, B., Kennedy, P., Seipel, B., Carlson, S. E., Biancarosa, G., & Davison, M. L. (2019). Can we learn from student mistakes in a reading comprehension assessment? *Journal of Educational Measurement, 56*, pp. 815 – 835. DOI.org/10.1111/jedm.12238
- McMaster, K. L., Espin, C. A., & van den Broek, P. (2014). Making connections: Linking cognitive psychology and intervention research to improve comprehension of struggling readers. *Learning Disabilities Research & Practice, 29*(1), 17-24.
- McMaster, K. L., van den Broek, P., Espin, C. A., White, M. J., Kendeou, P., Rapp, D. N., Bohn-Gettler, K., Carlson, S. E. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences, 22*, 100–111.
- 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. doi: 10.1080/10888430701530417