MOCCA’s Instructional Effects: More Than Just A Correct Answer

Mark L. Davison, University of Minnesota
Ben Seipel, University of Wisconsin, River Falls
Sarah Carlson and Gina Biancarosa, University of Oregon
Bowen Liu, University of Minnesota

This talk describes a test to measure instructional effects tracking three indicators: (1) reduction of student errors, (2) changes in the pattern of student errors, and (3) increases in efficiency (less time per correct response). It tracks response times and types of student mistakes. It also requires analysis of the response times and patterns of student mistakes.

Background

Outcomes of efficacy and effectiveness studies are often test scores, either a simple total score or an IRT score based on the items answered correctly. In this preliminary study using a measure of inferential reading comprehension, we considered two other outcome variables (a) the pattern of student errors and (b) student correct response times.

Prior think-aloud research suggests that poor comprehenders can be distinguished by reliance on two types of reading comprehension processes that are not sufficient for maintaining causal coherence (e.g., Coté, Goldman, & Saul, 1998; McMaster et al., 2012; Trabasso & Magliano, 1996a, 1996b; van den Broek et al., 2001; Wolfe & Goldman, 2005). The first is paraphrasing, or restating prior text. The second is lateral connecting, which focuses on making elaborations, associations, and self-explanations that are not causally coherent connections. In the test on which this research is based, items have three alternative responses: a correct (causally coherent inference) response, and two incorrect responses, a paraphrase and a lateral connection.

The test is computer administered so that both student responses and student response times are recorded. Automaticity theory, sometimes called efficiency or dual processing theory, suggests that as readers mature, the reading process becomes automatized and appropriately faster (Laberge & Samuels, 1974; Goldhammer, Naumann, Stelter, Tóth, Rölke, & Kleine, 2014; Perfetti, 1985; Skinner, Neddenriep, Bradley-Klug, 2002). This leads to the hypothesis that the rate of correctly responding to items may be an indication of reading efficiency.

Purpose

The purpose of this presentation is to compare more and less mature readers to determine whether they differ in more than just number correct, whether they also differ in the pattern of their errors and their response times when answering items correctly.

Setting

The participants were enrolled in elementary schools.

Population

Participants were in Grades 3-5 in the 2016 academic year. Samples sizes were 1577 3rd graders, 1499 4th graders, and 1216 5th graders in over 50 schools and 13 states.

Intervention

We are not comparing interventions. However, we will compare Grade 3-5 students’ test scores to evaluate whether more and less mature readers differ in types of errors and correct response times as well as number of correct responses. The study tentatively suggests that error patterns and response times may be useful outcome measures.

Research Design
**Test.** The test administered was the *Multiple-choice Online Causal Coherence Assessment* (MOCCA; Carlson et al., 2014). It contains 40 items each of which is a story with a missing sentence. The student must select one sentence out of three sentence response type options that best completes the story. The test is administered in a computer lab or classroom via a desktop computer, laptop, or tablet with no time limit. For progress monitoring, it has three forms per grade with stories written at the appropriate grade level. Within each grade, forms were randomly assigned.

**Sampling.** Students come from over 50 schools in 13 states. Test data have been collected, but demographic data are still being collected. Based on the district demographic statistics, we expect that Whites will be overrepresented. American Indian, Asian, Blacks, and Hispanics will be underrepresented as compared to the U.S. population, but there will be a substantial number of Hispanics.

**Data Collection**

MOCCA was administered from February 1 to June 16, 2016. Approximately 70% of students completed all 40 items. The mean total testing time was 35 minutes with a standard deviation of 15.

**Findings to Date**

Table 1 shows the mean, standard deviation, and internal consistency reliability (alpha) by grade and form for the causal coherent inference (correct), paraphrase, and lateral connect response scores, all of which can range from 1 – 40 since there are 40 items. Table 2 shows significant (*p* < .001) grade effects on the total correct and minutes per correct response. More mature 5th graders were both more accurate and faster.

We also ran a profile analysis (SPSS®, 1997; Davison & Davenport, 2002) on the incorrect response types: paraphrase, lateral connect, and no response (skip) (Figure 1). This analysis included three factors: Grade, Form within Grade, and Response Type. Both the Grade and Grade by Response Type effects were significant (*p* < .05). The Grade effect indicates that the more mature readers (4th and 5th graders) made fewer mistakes overall. The Grade by Response Type effect is the pattern effect. Paraphrase incorrect responses predominated for third graders. Fifth graders show far fewer omits than Paraphrases and Lateral Connects, possibly reflecting their greater efficiency. Together, the Grade effect and the Grade by Response Type effects suggest that less mature readers make more errors, but they also have a different pattern of errors marked by a preponderance of paraphrase responses.

**Conclusions Based on Findings to Date**

The study has several limitations. First, we do not have an intervention, and consequently we cannot say whether the observed effects are effects of intervention, natural development, or other experiences. MOCCA used grade-appropriate reading levels, so that different items were taken by the different grades. Furthermore, the approach requires tests like MOCCA that are designed to track types of student errors and item response times.

Nevertheless, the grade effect in our profile analysis suggests that the number of student errors was sensitive to educational, developmental, or experiential effects. The pattern of errors and the minutes-per-correct response were also sensitive to educational/developmental/experiential effects. While further research needs to disentangle educational/developmental/experiential effects, results suggest that this approach may show instructional effects in multiple ways: reduction in errors, change in error patterns, and response times. Procedures that combine these indicators (e.g. MANOVA) may also yield more sensitive and statistically more powerful tests of instructional effects.
References


Table 1.

Mean, Standard Deviation, and Internal Consistency Reliability (Alpha) for Each Response Type by Grade and Form

<table>
<thead>
<tr>
<th>Form</th>
<th>N</th>
<th>Causal</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Paraphrase</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Lateral Connect</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>α</td>
<td>M</td>
<td>SD</td>
<td>α</td>
<td>M</td>
<td>SD</td>
<td>α</td>
<td>M</td>
<td>SD</td>
<td>α</td>
</tr>
<tr>
<td>3.1</td>
<td>489</td>
<td>20.61</td>
<td>10.78</td>
<td>0.94</td>
<td>7.98</td>
<td>6.82</td>
<td>0.88</td>
<td>5.92</td>
<td>4.48</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>478</td>
<td>20.99</td>
<td>10.31</td>
<td>0.92</td>
<td>7.59</td>
<td>6.41</td>
<td>0.87</td>
<td>6.74</td>
<td>4.55</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>481</td>
<td>21.56</td>
<td>10.77</td>
<td>0.93</td>
<td>6.75</td>
<td>6.03</td>
<td>0.86</td>
<td>5.33</td>
<td>4.54</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>426</td>
<td>24.16</td>
<td>11.36</td>
<td>0.94</td>
<td>5.29</td>
<td>5.50</td>
<td>0.88</td>
<td>4.76</td>
<td>4.44</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>411</td>
<td>24.35</td>
<td>11.41</td>
<td>0.93</td>
<td>5.44</td>
<td>5.78</td>
<td>0.88</td>
<td>4.63</td>
<td>4.13</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>401</td>
<td>23.15</td>
<td>11.32</td>
<td>0.94</td>
<td>5.40</td>
<td>5.38</td>
<td>0.86</td>
<td>5.14</td>
<td>4.56</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>366</td>
<td>27.01</td>
<td>9.71</td>
<td>0.93</td>
<td>5.07</td>
<td>5.14</td>
<td>0.86</td>
<td>5.68</td>
<td>4.43</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.2</td>
<td>339</td>
<td>27.23</td>
<td>10.24</td>
<td>0.93</td>
<td>5.18</td>
<td>5.21</td>
<td>0.87</td>
<td>4.33</td>
<td>3.96</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>330</td>
<td>26.67</td>
<td>10.76</td>
<td>0.94</td>
<td>5.08</td>
<td>5.43</td>
<td>0.88</td>
<td>4.92</td>
<td>4.40</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>Total Correct Score (out of 40 Possible)</td>
<td>Minutes per Correct Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------</td>
<td>------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>Estimated Marginal Mean: 22.206</td>
<td>Standard Error: .262</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Estimated Marginal Mean: 2.010</td>
<td>Standard Error: .046</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>Estimated Marginal Mean: 25.492</td>
<td>Standard Error: .269</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Estimated Marginal Mean: 1.702</td>
<td>Standard Error: .048</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>Estimated Marginal Mean: 27.015</td>
<td>Standard Error: .299</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Estimated Marginal Mean: 1.650</td>
<td>Standard Error: .053</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Mean Number of Incorrect Responses by Response Type and Grade