Hammill Institute on Disabilities

Journal of Learning Disabilities
1–14

DOI: 10.1177/00222194221121340
journaloflearningdisabilities.sagepub.com
SAGE

# MOCCA-College: Preliminary Validity Evidence of a Cognitive Diagnostic Reading Comprehension Assessment

**Ben Seipel, PhD[1], Patrick C. Kennedy, PhD[2], Sarah E. Carlson, PhD[3], Virginia Clinton-Lisell, PhD[4], and Mark L. Davison, PhD[5]**

**Abstract**

As access to higher education increases, it is important to monitor students with special needs to facilitate the provision of appropriate resources and support. Although metrics such as ACT's (formerly American College Testing) "reading readiness" provide insight into how many students may need such resources, they do not specify *why* a student may need support or how to provide that support. Increasingly, students are bringing reading comprehension struggles to college. Multiple-choice Online Causal Comprehension Assessment-College (MOCCA-College) is a new diagnostic reading comprehension assessment designed to identify who is a poor comprehender and also diagnose *why* they are a poor comprehender. Using reliability coefficients, receiver–operating characteristic curve analysis, and correlations, this study reports findings from the first year of a 3-year study to validate the assessment with 988 postsecondary students who took MOCCA-College, a subset of whom also provided data on other reading assessments (i.e., ACT, $n = 377$; Scholastic Aptitude Test [SAT], $n = 192$; and Nelson–Denny Reading Test [NDRT], $n = 78$). Despite some limitations (e.g., the sample is predominantly females from 4-year institutions), results indicate that MOCCA-College has good internal reliability, and scores are correlated with other reading assessments. Through a series of analyses of variance (ANOVAs), we also report how students identified by MOCCA-College as good and poor comprehenders differ in terms of demographics, cognitive processes used while reading, overall comprehension ability, and scores on admissions tests. Findings are discussed in terms of using MOCCA-College to help gauge which students may be at risk of reading comprehension difficulties, identify why they may be struggling, and inform directions in actionable instructional changes based on comprehension processing data.

**Keywords**

reading comprehension, poor comprehenders, diagnostic assessments, college students

The percentage of postsecondary applicants who have the reading skills to succeed in college has decreased by nearly 10% in recent years (American College Testing [ACT], 2014, 2016). The ACT college readiness benchmark corresponds to a 50% chance of obtaining at least a B and a 75% chance of obtaining at least a C in a credit-bearing first-year college course and is equivalent to a 22 on the ACT Reading subtest (ACT, 2014, p. 24). This benchmark provides an indication of the likelihood of reasonable college success (i.e., timely degree completion; Allen & Radunzel, 2017). In both 2013 and 2014, only 44% of high school graduates who took the ACT met the benchmark for reading readiness, down from 52% in previous years (ACT, 2014). Additionally, reading readiness is especially low for *at-risk populations*, including first-generation students and those from ethnic and racial groups that are traditionally underserved in education. For example, reading readiness rates range from just 17% for African American students to 29% for Hispanic students (ACT, 2014). Low rates of college reading readiness are especially problematic because the ability to learn from discipline-specific text is a basic expectation of higher education (Holschuh, 2019). Thus, it is not surprising that reading skills and passing a developmental

[1]California State University, Chico, USA
[2]University of Oregon, Eugene, USA
[3]Georgia State University, Atlanta, USA
[4]University of North Dakota, Grand Forks, USA
[5]University of Minnesota, Minneapolis, USA

**Corresponding Author:**
Ben Seipel, California State University, Chico, 400 West First Street, Chico, CA 95929, USA.
Email: bseipel@csuchico.edu

reading course are major predictors of student retention after the first year of college (Bergey et al., 2017; Fike & Fike, 2008), and the need to identify students who struggle with reading and provide them with targeted support is clear (Biden, 2011; Holschuh, 2019; Poole, 2019).

The importance of reading, coupled with the prevalence of poor reading skills at the postsecondary level (ACT, 2014; Gorzycki et al., 2016), has led many institutions to offer reading support for students through developmental coursework, reading centers, and tutoring (Scott-Clayton & Rodriguez, 2014; Tinto, 2012). However, some readers at this level struggle not with basic reading skills (e.g., decoding) but with comprehension skills (e.g., generating appropriate inferences). This issue is explained by the *simple view of reading* in which reading comprehension is the product of two skills: decoding and language comprehension (Hoover & Gough, 1990). Decoding is the ability to determine the sounds of written words (Goff et al., 2005), and language comprehension is the ability to develop a mental representation of the text (Kintsch, 2019). Readers with adequate decoding skills, but who have difficulty understanding what they read, have been labeled *poor comprehenders* (Cain & Oakhill, 2006; Carlson et al., 2014; McMaster et al., 2012). Without a reliable and valid way to identify poor comprehenders, determine *which* aspects of comprehension are challenging for them, and assess changes in their reading comprehension skills, it is nearly impossible for institutions to know which mix of services are appropriate to offer. Moreover, despite a renewed push for the standardization of assessments that identify students in need of remediation (Boylan, 2009), such practices may be ineffective if they only provide evidence *that* students struggle without differentiating *why* they struggle, leading to inadequate or even inappropriate reading support.

One leading theory in reading comprehension research that has provided a foundation to understand why readers struggle with comprehension is the *construction-integration* (CI) *model of comprehension* (Kintsch, 2019), which articulates three levels of representation involved in developing mental representations of text. The first level is the surface structure, which consists of the literal words and syntax in which decoding skills are necessary. Second, information in the surface structure is connected within the text to develop the textbase level of representation. Third, a situation model level of representation is developed when readers connect their background knowledge to the textbase, thus filling in gaps in the text with necessary background knowledge (Kintsch & van Dijk, 1978).

The field has gained valuable insight into higher-level cognitive skills to determine *how* and *why* some readers struggle with comprehension at the textbase and situation model levels (Cain & Oakhill, 1999, 2006). For example, previous think-aloud research (i.e., recorded verbal thought responses or processes while reading) with children has found that some readers tend to focus heavily on repeating or summarizing individual sentences (i.e., *paraphrasing*) without considering ideas across a text, while other students focus more on connecting to their background knowledge (i.e., *elaborating* about information related but also irrelevant to the text) rather than honing in on the key ideas in the text (i.e., making causal connections; McMaster et al., 2012). Although paraphrasing and elaborating are helpful cognitive skills, they are not sufficient for comprehension, and poor comprehenders have been shown to overly rely on these processes to the detriment of generating causally coherent inferences (i.e., determining why something occurs in a text)—which are required for comprehension. However, traditional reading comprehension assessments do not generally assess the underlying processes that readers use during reading, nor do they identify why readers struggle during comprehension (Magliano et al., 2007). Instead, they focus on the result (i.e., product) of reading after reading is complete (e.g., what is recalled or remembered).

Although recent research on the comprehension processes children use while reading has helped push the field into new directions (e.g., classroom interventions; McMaster et al., 2012; computer-based interventions: ELLCII and TeLCI; McMaster et al., 2019), much less is known about the comprehension processes of older readers (i.e., college students), particularly those who struggle with reading comprehension and are not considered *reading ready* for college success (ACT, 2014, 2016). Beyond college admissions tests, there are few reading comprehension assessments intended for use with postsecondary readers. Moreover, those that are available are typically not designed to identify why those students struggle with reading. Thus, there exists a clear need for new assessments that can differentiate comprehension skills of postsecondary students. In addition, there is a particular need for diagnostic assessments that identify the processes that postsecondary students use during reading to better understand their reading skills and strategies (e.g., McNamara, 2017; McNamara & Kendeou, 2011) and provide scores that are informative for both postsecondary students and personnel who teach developmental or other supportive reading courses. Students and personnel could use such diagnostic data to guide decisions for instructional support such as individualized tutoring (Colver & Fry, 2016), computerized questioning (Magliano et al., 2011), intelligent tutoring (McMaster et al., 2019), or game-based learning (Ronimus et al., 2019) needed to help improve reading comprehension strategies and processes generated during reading. To begin to address these needs, the current study reports on the initial development of and preliminary validation evidence for Multiple-choice Online Causal Comprehension Assessment-College (MOCCA-College), a diagnostic assessment of reading comprehension that differentiates the reading comprehension processes of postsecondary students.

## Purposes of the Present Study

This article reports data from the first year of a 3-year measurement development study and has two primary purposes:

1. To report preliminary convergent and discriminant validity evidence regarding the use of MOCCA-College as an assessment of reading comprehension for postsecondary students; and
2. To evaluate construct validity evidence for MOCCA-College as a diagnostic measure of reading comprehension by investigating differences between students identified as good and poor comprehenders with respect to demographic characteristics, comprehension processes, and comprehension efficiency.

To do so, we report pilot data from a demographically diverse convenience sample of college students, including performance on MOCCA-College, available data from standardized admissions assessments (i.e., Scholastic Aptitude Test [SAT] and ACT), and reading comprehension and reading rate scores from the Nelson-Denny Reading Test.

To address our first purpose, we evaluated convergent validity for MOCCA-College using correlations between the MOCCA-College number correct score and reading scores on the various standardized assessments. Discriminant validity was evaluated using correlations between the MOCCA-College number correct score and standardized mathematics scores. Based on the construct similarity between the measures, we anticipated reasonably strong positive correlations between the MOCCA-College number correct score and standardized reading scores, and negative correlations between standardized reading scores and the MOCCA-College distractor subscores (see more about MOCCA-College subscores under *Measures*). Given that mathematics assessments typically include word problems in which students need to combine the situation model from the question with the necessary mathematical skills to solve the problem (i.e., the problem model; Nathan et al., 1992), we anticipated that correlations between the MOCCA-College correct score and mathematics scores would also be positive, but hypothesized that they would be weaker than those between MOCCA-College and the reading scores.

To address our second purpose, we tested for differences between comprehender type on measures of general reading comprehension, reading rate, and math performance. Based on findings from the original MOCCA developed for late elementary readers, we expected to find similar patterns of response trends between good and poor comprehenders such that differences between groups would be more pronounced for measures of reading than for measures of math.

## Method

### Participants

Participants were recruited from three medium-to-large, regional, comprehensive universities located in geographically distinct regions of the United States (i.e., the West, Midwest, and Southeast). Participants were drawn from a combination of First-Year Experience courses; TRIO programs (i.e., federally supported programs that provide services to students from underserved, disadvantaged, and low-income backgrounds); Educational Opportunity Programs (i.e., "Summer Bridge"); psychology and educational psychology participant pools; and psychology, English, and education courses. In total, a demographically diverse convenience sample of 1,160 postsecondary students consented to participate in the study.

Fifty participants completed fewer than 10 of the 50 MOCCA-College items and were excluded from subsequent analysis due to insufficient data. An additional 122 participants were excluded because they declined to provide demographic data. Scores on admissions data did not differ between excluded and included participants (see Note 1). The final analytic sample consisted of 988 participants who completed at least 10 MOCCA-College items and provided demographic data. Of these 988 participants, ACT data were available for 377, and SAT data were available for 192. In addition, a subset of 85 participants from two of the participating institutions completed an additional measure of reading comprehension and reading rate through in-person data collection, 78 of whom provided demographic data and were included in the final sample. Participants were not excluded from the study due to a documented learning disability (LD), but information about the number of students with LD was not available.

### Measures

*MOCCA-College.* The MOCCA-College is a discourse-level maze task designed to identify postsecondary students who struggle with causal comprehension. It is based on an earlier version of MOCCA developed for readers in Grades 3 through 5 (Biancarosa et al., 2016; Carlson et al., 2014; Davison et al., 2018; Liu et al., 2019; see Figure 1). Each of the four MOCCA-College test forms consists of 50 items, which are short, seven-sentence narrative or expository texts, written within a specific Flesch-Kincaid Grade Level (FKGL) range for the target population (FKGL ~6.0–14.0, $M$ = 8.74, $SD$ = 2.50; Kincaid et al., 1975). Unlike a traditional maze task in which items are generated by deleting

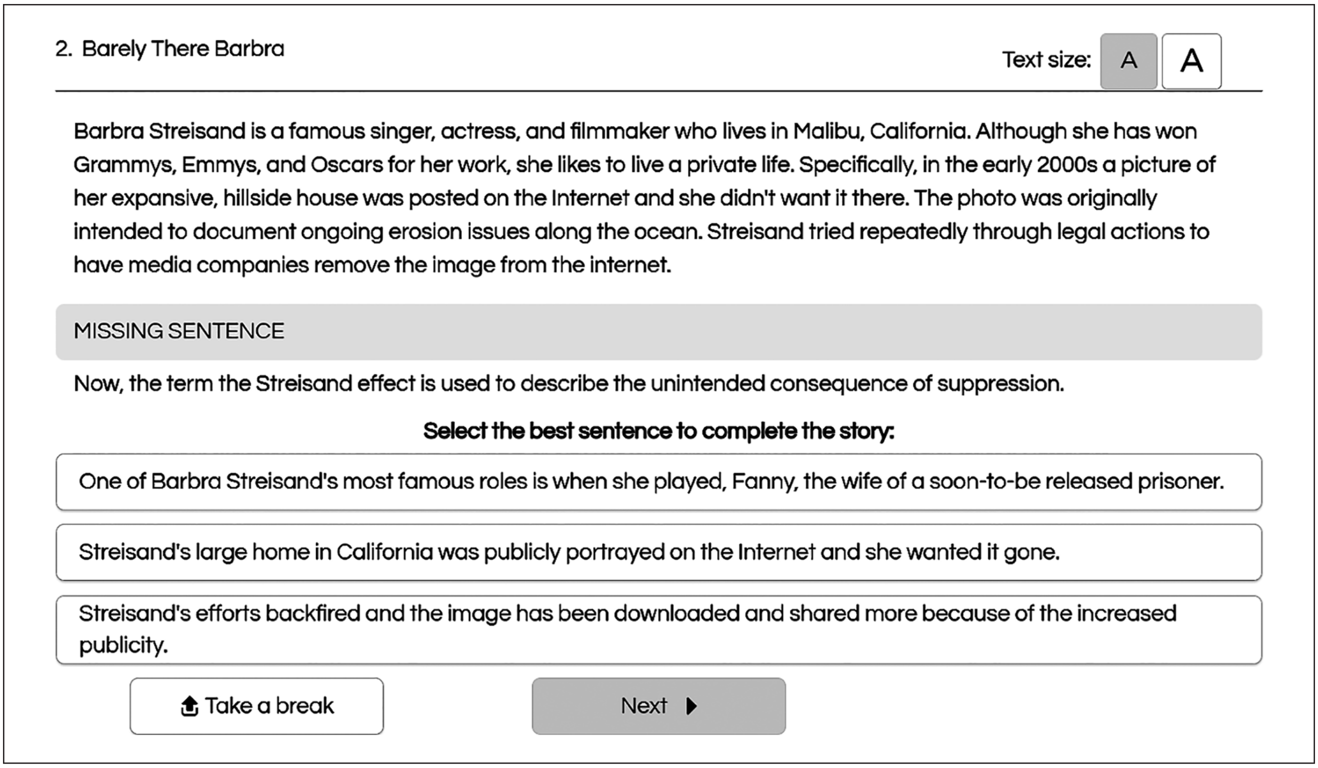


**Figure 1.** Screenshot of a MOCCA-College demonstration item.
*Note.* This figure illustrates a demonstration item with line "MISSING SENTENCE" embedded in the text. Examinees select their choice by clicking one of the sentences below the main text. In this example, the first response is an elaboration, the second response is a summary/paraphrase of the main idea, and the third response causally completes the text. MOCCA-College = Multiple-choice Online Causal Comprehension Assessment-College.

every *n*th word, MOCCA-College deletes the entire sixth sentence from the text. Examinees read each item and replace the missing sentence by choosing from among the three multiple-choice response types: (a) *causally coherent inferences* (CCI), (b) *paraphrases* (PAR), and (c) *elaborations* (ELA). CCIs are keyed as correct and represent the response that best completes the text in a causally coherent manner. MOCCA-College is thus a distractor-driven assessment (Sadler, 1998) that distinguishes not only whether readers selected the correct answer but also *why* they may not have selected the correct answer based on which of the meaningful distractors was chosen.

The incorrect PAR response reiterates a previous sentence. This mimics what one group of poor intermediate comprehenders tended to do when they failed to make causally coherent inferences (e.g., Carlson et al., 2014; McMaster et al., 2012). The incorrect ELA response involves a connection to the text based on background knowledge that may be tangential. This response mimics what another group of poor intermediate comprehenders tended to do when they failed to make causally coherent inferences (e.g., Carlson et al., 2014; McMaster et al., 2012). Note that readers may draw on background knowledge to make either causally coherent inferences or elaborations. What distinguishes the two in MOCCA-College is whether the inference closes the causal chain of events in the text. When a participant completes MOCCA-College, they receive scores representing the number of times they selected each of the three response types, and a percent correct score corresponding to the percentage of items attempted for which they selected the CCI.

MOCCA-College is administered completely online via a secure, individualized login. The assessment is self-paced but takes approximately 1 hr to complete in its entirety, including instructions and practice items. Because one of the purposes of the project for which these data were collected was to evaluate the performance of the newly developed MOCCA-College items, no ceiling rule to discontinue testing was used. However, as detailed under *Participants*, examinees who answered fewer than 10 items were excluded from the analyses presented here.

*Demographic survey.* A demographic survey was administered via Qualtrics (2019). The survey was hyperlinked at the end of the MOCCA-College assessment. The demographic survey included items regarding age, gender, sex, race, ethnicity, primary language, student status, employment status, the highest level of parent education, and nontraditional student identifiers (i.e., transfer, returning, marital status, parent, primary caregiver, primarily online, and veteran). The demographic survey took approximately 4 min to complete. Participants could skip any item or select "I prefer not to answer."

*Admissions data.* As available, admissions test score data (i.e., ACT and SAT scores) were collected from participating institutions to evaluate the convergent and discriminant validity of MOCCA-College test scores. The ACT is a multiple-choice college readiness test designed to measure skills that are typically acquired in secondary education and are useful in postsecondary education (ACT, 2014). The ACT provides a composite score and four subscores in English, Mathematics, Reading, and Science, all of which have high reported reliability correlation coefficients ($\alpha$ = .97, .92, .90, .88, and .85 respectively) for college-bound students (ACT, 2014). Results reported here focus on the composite and Reading and Mathematics subtests. Similarly, the SAT is a college readiness test that provides test scores in three areas: Reading, Writing and Language, and Mathematics. Most items are multiple-choice, but the test also includes some sentence completion and constructed response items. Reported reliability correlation coefficients for the SAT scores are high: Reading range = .91–.93, Mathematics range = .92–.94, and Writing and Language range = .88–.93 (College Board, 2012). In addition, the SAT provides two scaled section scores that range from 200 to 800 for (a) Evidence-Based Reading and Writing and (b) Math. All analyses involving the SAT use section scores. The availability of admissions data was somewhat limited due, in part, to the evolving nature of *test optional* or *test flexible* admission requirements by many institutions of higher education (e.g., Buckley, 2020). Specifically, one of the three participating institutions did not require college-admission test scores. In addition, not all participants had test scores on file. For example, transfer students with associates' degrees are not required to submit ACT or SAT scores at most 4-year institutions, including those in the current study.

*Nelson–Denny Reading Test (NDRT).* The NDRT is a traditional paper-and-pencil test of reading comprehension, reading rate, and vocabulary for high school and adult readers. The NDRT comprehension and reading rate subtests were included in this study as measures of concurrent reading validity. The NDRT requires examinees to read passages and answer multiple-choice questions. Passages are also used to measure reading rate. The NDRT provides raw, scaled, and normative percentile rank scores; NDRT standardized comprehension subscores have a mean of 100 and a standard deviation of 15. The authors report internal consistency reliability coefficients ranging from .85 to .95, alternate-form reliability coefficients between .88 and .94, and test–retest reliability coefficients between .89 and .98 (Fishco, 2018; Forms I and J).

## Procedure

First, participants were recruited through a combination of university courses, psychology and education course recruitment systems (e.g., Sona Systems), and social media. Second, after informed consent, participants completed one of the four randomly assigned MOCCA-College forms. Third, immediately upon finishing the assessment, participants were asked to complete the brief, optional, demographic survey. All participants who completed the assessment were compensated with either course credit or a gift card to a national retailer. Fourth, a subset of participants ($n$ = 85) were also recruited and completed the NDRT in person with trained research staff, 78 of whom provided demographic data and are included in the analytic sample. These additional measures took approximately 45 min to complete, and individuals who did so were compensated with a gift card to an online retailer. Fifth, admissions data (i.e., ACT [$n$ = 377] and SAT [$n$ = 192]) were requested from and provided by participating institutions for analyses.

## Analyses

Our first analytic goal was to identify a preliminary performance threshold, or cut score, on MOCCA-College that could be used to categorize performance on MOCCA-College as indicative of either good comprehension or poor comprehension. Importantly, the goal of these analyses was not to establish a threshold that would be shared with end users but to facilitate comparisons across comprehension types using data from all participants who took MOCCA-College, rather than limiting those analyses to just the smaller subset of participants for whom other reading achievement data were available. To identify this preliminary threshold, we used receiver–operating characteristic (ROC) curve analyses to evaluate the predictive relationship between a range of MOCCA-College percent correct scores and criterion-level performance on all other available measures of reading achievement. The ACT and SAT criteria were set as the published college and career readiness benchmark scores for each test (i.e., 22 for the ACT, 2021; and 480 for the SAT, College Board, 2021). The criterion for the NDRT was set as the 40th percentile on the total reading score (Fishco, 2018).

Classification accuracy of the MOCCA-College percent correct score was evaluated using the area under the curve (AUC) statistic. In traditional ROC curve analyses, the AUC provides a measure of how well a screener predicts outcomes on the criterion measure. Because MOCCA-College is not designed as a screener for these outcomes per se, we interpreted these analyses as an indicator of the extent to which performance on MOCCA-College is indicative of reading comprehension generally. For each outcome, we identified the MOCCA-College percent correct score that best balanced sensitivity (i.e., how well a given score correctly identifies students who have not met the criterion) and specificity (i.e., how well that same score correctly identifies those students who have met the criterion) thresholds.

Demographic and MOCCA-College subscore data were then summarized descriptively, both overall, and by comprehender type. To evaluate validity evidence for the intended uses of MOCCA-College, we correlated MOCCA-College subscores with other available measures. Evidence for construct validity was based on the joint evaluation of convergent validity correlations between the MOCCA-College subscores, the NDRT ($n$ = 78), and the ACT ($n$ = 377) and SAT ($n$ = 172) reading subtests, and discriminant correlations between the MOCCA-College subscores and the ACT and SAT math subtests. The primary correlations of interest (i.e., those involving MOCCA-College) with the smallest sample size (NDRT; $n$ = 78) were powered at .80 to detect a two-tailed correlation as small as .31 (Weathington et al., 2012). Finally, to investigate differences between students identified by MOCCA-College as good and poor comprehenders, we conducted a series of analyses of variance (ANOVAs; between-subjects factor: comprehension ability as defined by MOCCA-College percent correct) on MOCCA-College subscores, admissions data, and scores on the NDRT, and reported effect sizes for each comparison using Hedges' $g$ (Hedges, 1981).

## Results

### ROC Curve Analyses

Sample sizes for the ROC curve analyses varied by outcome (i.e., ACT $n$ = 377, SAT $n$ = 192, NDRT $n$ = 78). AUCs for the MOCCA-College percentage correct score predicting the ACT, SAT, and NDRT criterion scores were .72, .72, and .66, respectively, all values at the low end of the adequate range (Hosmer et al., 2013). Comparing these AUC values to commonly cited efficacy criteria (e.g., National Center on Intensive Intervention [NCII], 2020; Swets, 1992), the MOCCA-College percentage correct score should not be used to predict performance on these measures. However, the MOCCA-College percentage correct scores that best balanced sensitivity and specificity were 84%, 82%, and 82% for the ACT, SAT, and NDRT, respectively. This similarity in cut scores across different reading outcomes does suggest a reasonably strong relationship between performance on MOCCA-College and a range of reading comprehension outcomes, which we believe justifies the use of these analyses as a guide for choosing a threshold on MOCCA-College for the purposes of making preliminary comparisons between comprehender type. Based on these results, we selected a percentage correct score of 82% as the threshold for subsequent comparisons and applied that threshold to the percentage correct scores for all study participants, resulting in 638 students who exhibited good comprehension of MOCCA-College, and 350 students who exhibited poor comprehension on MOCCA-College.

### Demographic Survey Results

To measure potential test biases and help characterize differences in performance, all participants were asked to voluntarily complete a demographic survey at the end of the assessment; 988 students did so. Details regarding overall demographics as well as a breakdown by good and poor comprehender groups are reported in Table 1. The average age of participants who completed the survey was 22.56 years ($SD$ = 6.78). More than two-thirds of participants (69%) identified as female, 29% identified as male, and 1% identified as genderfluid, transgender, or gender nonconforming. In terms of racial identity, 62% of respondents identified as White, 13% identified as Black or African American, 6% identified as Asian, 2% identified as Native American, Alaska Native, Hawaiian, or Pacific Islander, and 5% reported a different racial identity. In terms of ethnic identity, 17% reported being Latino/a/x. The primary language of 91% of participants was English, with 3% reporting Spanish, and 2% reporting a different language.

These demographics are similar to those reported for the national average of college students in the United States for race and ethnicity (54.7% non-Hispanic White; 14% Black or African American; 8.4% Asian; 3.1% other; 19.4% Hispanic; U.S. Census Bureau, 2018). Notably, our sample includes a higher percentage of female students than the general U.S. college student population (55.5% female; Hanson, 2021), and all participants attend public 4-year intuitions. In the general postsecondary population, approximately 35% of students attend private institutions and nearly a quarter of students attend 2-year institutions. The breakdown by good and poor comprehender groups closely paralleled the overall data, although students identified as poor comprehenders were somewhat more likely to be people of color and speak a primary language other than English than were students identified as good comprehenders (67% of good comprehenders identified as White and 95% spoke English as their primary language, compared with 52% and 84% of poor comprehenders, respectively).

The average age of the 78 participants who completed the NDRT was 21.14 ($SD$ = 4.66). Nearly two-thirds (65%) identified as female, 33% identified as male, and 1% identified as genderfluid, transgender, or gender nonconforming. In terms of racial identity, 81% of these respondents identified as White, 4% identified as Black or African American, 4% identified as Asian, 1% identified as Native American, Alaska Native, Hawaiian, or Pacific Islander, and 3% reported a

**Table 1.** Demographics for the Full Sample and by Comprehender Subgroup.

| | Full sample | | Good comprehender | | Poor comprehender | |
|---|---|---|---|---|---|---|
| Demographic characteristic | *N* | % | *n* | % | *n* | % |
| All respondents | 988 | 100 | 638 | 64 | 350 | 35 |
| Gender/sex | | | | | | |
| Female | 684 | 69 | 450 | 71 | 234 | 67 |
| Male | 285 | 29 | 176 | 28 | 109 | 31 |
| Genderfluid/transgender/nonconforming | 13 | 1 | 9 | 1 | 4 | 1 |
| Race | | | | | | |
| White | 680 | 62 | 477 | 67 | 203 | 52 |
| Black/African American | 139 | 13 | 80 | 11 | 59 | 15 |
| Asian | 71 | 6 | 39 | 6 | 32 | 8 |
| American Indian/Alaska Native/Pacific Islander/Hawaiian | 23 | 2 | 15 | 2 | 8 | 2 |
| Other | 59 | 5 | 32 | 5 | 27 | 7 |
| Ethnicity: Latino/a/x | 168 | 17 | 94 | 15 | 74 | 21 |
| Primary language | | | | | | |
| English | 919 | 91 | 617 | 95 | 302 | 84 |
| Spanish | 31 | 3 | 7 | 1 | 24 | 7 |
| Other | 21 | 2 | 9 | 1 | 12 | 3 |
| Student status | | | | | | |
| Full time | 925 | 93 | 592 | 93 | 333 | 95 |
| Freshman | 258 | 26 | 149 | 23 | 109 | 31 |
| Sophomore | 189 | 19 | 127 | 20 | 62 | 18 |
| Junior | 184 | 18 | 116 | 18 | 68 | 19 |
| Senior | 178 | 18 | 109 | 17 | 69 | 20 |
| Other | 167 | 17 | 131 | 21 | 36 | 10 |
| Employment status | | | | | | |
| Regular work schedule | 520 | 53 | 338 | 53 | 182 | 52 |
| Almost never work | 216 | 22 | 148 | 23 | 68 | 20 |
| No regular work schedule | 215 | 22 | 140 | 22 | 75 | 22 |
| Highest parental education level | | | | | | |
| Less than high school | 63 | 6 | 26 | 4 | 37 | 11 |
| Secondary (high school diploma or equiv.) | 169 | 17 | 100 | 16 | 69 | 20 |
| Vocational school | 27 | 3 | 19 | 3 | 8 | 2 |
| Some college, no degree | 145 | 15 | 83 | 13 | 62 | 18 |
| Associate's degree | 89 | 9 | 65 | 10 | 24 | 7 |
| Bachelor's degree | 269 | 27 | 193 | 30 | 76 | 22 |
| Master's degree | 145 | 15 | 103 | 16 | 42 | 12 |
| Terminal degree | 58 | 6 | 41 | 6 | 17 | 5 |
| I don't know | 13 | 1 | 6 | 1 | 7 | 2 |
| Nontraditional student identifier | | | | | | |
| Transfer | 144 | 12 | 92 | 11 | 52 | 13 |
| Returning to school | 110 | 9 | 77 | 10 | 33 | 8 |
| Married/domestic partnership | 89 | 7 | 70 | 9 | 19 | 5 |
| Primarily online | 82 | 7 | 65 | 8 | 17 | 4 |
| Parent | 66 | 6 | 48 | 6 | 18 | 5 |
| Primary caregiver | 23 | 2 | 20 | 2 | 3 | 1 |
| Veteran | 18 | 2 | 15 | 2 | 3 | 1 |

*Note*. Percentages do not always sum to 100 due to rounding and incomplete responses at the item level.

different racial identity. In terms of ethnic identity, 17% reported being Latino/a/x. The primary language of 90% of validating task participants was English, with 2% reporting Spanish and 2% reporting a different language.

**Table 2.** Descriptive Statistics for MOCCA-College for the Full Sample and by Comprehender Subgroup.

| | Full sample[a] | | | Good comprehenders[b] | | | Poor comprehenders[c] | | |
|---|---|---|---|---|---|---|---|---|---|
| Measure | *M* | *SD* | Range | *M* | *SD* | Range | *M* | *SD* | Range |
| CCI | 36.94 | 11.87 | 2–50 | 42.39 | 9.00 | 10–50 | 26.99 | 9.84 | 2–40 |
| PAR | 4.55 | 5.10 | 0–28 | 1.80 | 1.68 | 0–8 | 9.57 | 5.42 | 0–28 |
| ELA | 4.30 | 4.35 | 0–23 | 2.16 | 1.56 | 0–7 | 8.21 | 5.04 | 0–23 |
| Total testing time (in min) | 45.32 | 30.31 | 2.54–231.38 | 47.19 | 28.88 | 11.26–231.38 | 41.92 | 32.53 | 2.54–217.22 |
| Reading efficiency[d] | 1.34 | 1.05 | 0.16–12.41 | 1.14 | 0.65 | 0.28–5.51 | 1.70 | 1.46 | 0.16–12.41 |

*Note*. CCI = casually coherent inferences; PAR = paraphrases; ELA = elaborations; MOCCA-College = Multiple-choice Online Causal Comprehension Assessment-College.
[a]$N$ = 988. [b] $n$ = 638. [c] $n$ = 350. [d] Minutes per correct answer.

## MOCCA-College Reliability and Descriptives

The average difficulty of all items across all forms was $\theta$ = .81 ($\theta$ = .81, .78, .80. and .83 for Forms 1 through 4, respectively; Davison et al., 2020). The average point-biserial correlation across all items across all forms was $r_{pb}$ = .80 ($r_{pb}$ = .80, .77, .78, and .83 for Forms 1 through 4, respectively), which is considered good (Varma, 2006). Overall reliability for all four forms was excellent ($\alpha$ = .92–.95; Davison et al., 2020). Similarly, reliabilities of the paraphrase and elaboration subscores across forms were both very good (paraphrase $\alpha$ = .86–.88, elaboration $\alpha$ = .77–.88). Descriptive statistics for MOCCA-College are presented in Table 2 for the CCI total score (i.e., the number of items answered correctly), the PAR and ELA distractor subscores, total testing time, and reading efficiency. Each measure is reported both for the full sample and disaggregated by students identified by MOCCA-College as good and poor comprehenders. By definition, the 638 participants identified as good comprehenders chose more CCI responses and fewer PAR and ELA responses than did the 350 participants identified as poor comprehenders. Good comprehenders also spent an average of slightly more than 5 additional min on the test (Hedges' $g$ = .17, a moderately small effect) but less time per correct response (Hedges' $g$ = .55, a moderately large effect). Finally, participants identified as good comprehenders completed slightly more items ($M$ = 46.50, $SD$ = 9.15) than did participants identified as poor comprehenders ($M$ = 44.96, $SD$ = 11.27), $F(986)$ = 5.43, $p < .05$; Hedges' $g$ = .19, a moderately small effect).

## Validation Analyses

Descriptive statistics for the available ACT, SAT, and NDRT scores are reported in Table 3, including scores disaggregated by comprehender type. On average, students identified by MOCCA-College as poor comprehenders had lower scores on all measures except reading rate, a measure for which higher scores represent less fluent reading. Correlations between MOCCA-College subscores and the available validation measures are reported in Table 4. As predicted, CCI scores were positively and moderately, but significantly, correlated with scores from measures representing convergent validity, including the ACT Reading ($r$ = .43, $n$ = 377), SAT Reading and Writing ($r$ = .37, $n$ = 192), and the NDRT ($r$ = .32, $n$ = 78). In addition, as predicted correlations between CCI scores and measures of discriminant validity, including math subscores were of a similar magnitude but slightly lower than those of the reading scores (i.e., ACT Math $r$ = .31, $n$ = 377; SAT Math $r$ = .43, $n$ = 192). Finally, as predicted, MOCCA-College distractor (PAR, ELA) scores were negatively and moderately correlated with standardized measures of reading and math. Specifically, these correlations tend to be more strongly negative for standardized measures of reading than for standardized measures of math, providing further evidence of construct validity.

## Good/Poor Comprehender Differences

To investigate potential differences between students identified by MOCCA-College as good ($n$ = 638) and poor ($n$ = 350) comprehenders, we conducted a series of ANOVAs on the MOCCA-College reading comprehension efficiency score, admissions data, and other measures of convergent and discriminant validity.

*MOCCA-College reading comprehension efficiency.* An ANOVA testing the impact of participants identified as good vs. poor comprehenders on minutes per correct response showed that good comprehenders were more efficient ($M$ = 1.14 minutes per correct item; $SD$ = .65) than poor comprehenders ($M$ = 1.70 minutes per correct item; $SD$ = 1.46), a statistically significant difference, $F(1, 986)$ = 68.4; $p < .001$; Hedges' $g$ = .53, a moderately large difference.

### Admissions data

*ACT.* An ANOVA testing the association between comprehension ability and ACT Composite scores showed a moderate, statistically significant difference, $F(1, 375)$ = 67.2,

**Table 3.** Descriptive Statistics for Validation Measures for the Full Sample and by Comprehender Subgroup.

| | Full sample | | | | Good comprehenders | | | | Poor comprehenders | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Measure | *N* | *M* | *SD* | Range | *n* | *M* | *SD* | Range | *n* | *M* | *SD* | Range | Hedges' *g* |
| ACT test | | | | | | | | | | | | | |
| Composite | 377 | 25.26 | 4.05 | 16–35 | 253 | 26.37 | 3.90 | 17–35 | 124 | 23.01 | 3.37 | 16–32 | .90 |
| Reading | 377 | 26.19 | 4.98 | 13–36 | 253 | 27.54 | 4.72 | 15–36 | 124 | 23.43 | 4.33 | 13–35 | .89 |
| Math | 377 | 24.43 | 4.26 | 15–36 | 253 | 25.14 | 4.25 | 15–36 | 124 | 22.97 | 3.91 | 16–31 | .52 |
| SAT test | | | | | | | | | | | | | |
| Total | 192 | 1,027.60 | 135.08 | 750–1,530 | 115 | 1,073.22 | 135.16 | 810–1,530 | 77 | 959.48 | 106.82 | 750–1,220 | .91 |
| Read/write | 192 | 517.34 | 68.21 | 350–800 | 115 | 539.57 | 66.23 | 370–800 | 77 | 484.16 | 56.99 | 350–600 | .88 |
| Math | 192 | 510.94 | 79.34 | 330–800 | 115 | 533.04 | 80.08 | 350–800 | 77 | 477.92 | 66.02 | 330–640 | .73 |
| NDRT | | | | | | | | | | | | | |
| Total | 78 | 113.01 | 15.16 | 70–143 | 47 | 116.34 | 15.00 | 70–143 | 31 | 107.97 | 14.19 | 72–135 | .57 |
| Rate | 78 | 253.96 | 84.20 | 82–566 | 47 | 259.70 | 78.79 | 130–566 | 31 | 245.26 | 92.47 | 82–515 | .17 |

*Note.* ACT = American College Testing; NDRT = Nelson-Denny Reading Test; NDRT Total = NDRT total comprehension score; NDRT Rate = NDRT reading rate; MOCCA-College = Multiple-choice Online Causal Comprehension Assessment-College.

**Table 4.** Correlations Between Scores on MOCCA-College and Validation Measures.

| Score | *n* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. MOCCA-C CCI | 988 | — | | | | | | | | | |
| 2. MOCCA-C PAR | 988 | −.52** | — | | | | | | | | |
| 3. MOCCA-C ELA | 988 | −.48** | .74** | — | | | | | | | |
| 4. ACT Composite | 377 | .47** | −.44** | −.36** | — | | | | | | |
| 5. ACT Reading | 377 | .43** | −.42** | −.36** | .86** | — | | | | | |
| 6. ACT Math | 377 | .31** | −.30** | −.22** | .82** | .53** | — | | | | |
| 7. SAT Total | 192 | .44** | −.28** | −.19* | —[a] | —[a] | —[a] | — | | | |
| 8. SAT Read/write | 192 | .37** | −.30** | −.16* | —[a] | —[a] | —[a] | .89** | — | | |
| 9. SAT Math | 192 | .43** | −.20* | −.18* | —[a] | —[a] | —[a] | .91** | .64** | — | |
| 10. NDRT total | 78 | .32** | −.21 | −.08 | .51**[b] | .63**[b] | .10[b] | .46[c] | .59*[c] | .22[c] | — |

*Note.* MOCCAC = Multiple-choice Online Causal Comprehension Assessment-College; CCI = causally correct inference; PAR = paraphrase; ELA = elaboration; NDRT = Nelson-Denny Reading Test; MOCCA-College = Multiple-choice Online Causal Comprehension Assessment-College.
*n* = number of observations per variable, which corresponds to the pairwise sample size except where noted.
[a]No correlation reported, *n* = 4. [b] *n* = 42. [c] *n* = 17.
**p* < .05. ***p* < .01.

$p < .001$, with the good versus poor comprehender distinction explaining about 15% of the variance in scores ($\eta^2$ =.15, Hedges' $g$ = .90, a large effect). Similarly, a test of the association between comprehension ability and ACT Reading scores showed a statistically significant difference, $F(1, 375) = 66.7$, $p < .001$, that explained about 15% of the variance in scores ($\eta^2 = .15$, Hedges' $g = .89$, a large effect). The two groups showed a smaller, but still statistically significant difference on ACT Math subscores, $F(1, 375) = 25.5$, $p < .001$, $\eta^2 = .06$; Hedges' $g = .52$, a moderate effect.

*SAT.* An ANOVA testing the impact of comprehension ability on SAT Total scores showed a moderate, statistically significant difference, $F(1, 190) = 39.2$, $p < .001$, explaining about 17% of the variance in scores ($\eta^2 = .17$, Hedges' $g = .91$, a large effect). A test of comprehension ability on SAT Reading and Writing subscores also showed a moderate and statistically significant difference, $F(1, 190) = 36$, $p < .001$, which explained about 16% of the variance in scores ($\eta^2 = .16$, Hedges' $g = .88$, a large effect). An ANOVA testing the impact of comprehension ability on SAT Math scores also showed a statistically significant difference, $F(1, 190) = 25.1$, $p < .001$, explaining about 12% of the variance in scores ($\eta^2 = .12$, Hedges' $g = .73$, a large effect).

*Nelson-Denny Reading Test (NDRT).* An ANOVA testing the association of good vs. poor comprehenders with NDRT Total scores showed a small, but significant difference, $F(1, 76) = 6.07$, $p = .02$, explaining about 7% of the variance in scores ($\eta^2 = .07$, Hedges' $g = .57$, a moderate effect). An ANOVA of reading rate showed a small, non-significant difference, $F(1, 76) = .55$, $p = .46$, explaining <1% of the variance ($\eta^2 < .01$, Hedges' $g = .17$, a small effect).

## Discussion

Findings from this study are discussed in the context of our purposes: (a) to report preliminary validity evidence for MOCCA-College and (b) to identify differences between students identified by MOCCA-College as good and poor comprehenders. The results are also discussed for how best to design and use cognitive diagnostic assessments to inform tutoring, instruction, placements, or individual reading comprehension practice. Finally, we discuss the limitations of the study and future directions.

### *Validity Evidence*

The current study provides preliminary validity evidence for MOCCA-College as a measure of causal reading comprehension that could be used to distinguish between good and poor comprehenders and determine *why* they struggle with comprehension processing. On admissions and validation measures, scores on MOCCA-College were more highly correlated with measures of reading comprehension (i.e., NDRT) than they were with ACT math but not SAT math. These findings parallel, in part, results from the elementary version of MOCCA, which correlated more highly with other English language arts or reading tests than with math tests (Biancarosa et al., 2019; Davison et al., 2019). Notably, it is common for math and reading scores to be moderately correlated, as they are in this study (e.g., Davis et al., 2014; Dorans, 1999; Henry et al., 2014) given the overlap in skills in the two domains (e.g., mathematical story problems that require reading skills to be solved). Importantly, however, correlations with MOCCA-College tended to be higher for reading scores than for mathematics scores, which aligned with our predictions.

Although these findings represent important preliminary support for the convergent (reading) validity for MOCCA-College, evidence of discriminant (math) validity of MOCCA-College is mixed. One reason for these mixed results could be due to the fact that math, in general, becomes more difficult for students preparing for college than younger students. In addition, ACT and SAT Math may measure different aspects of math college readiness. However, this would need to be further examined in future research. Thus, although preliminary, these different types of validity show that performance on MOCCA-College may be predictive of scores on other measures that are also predictive of college success (i.e., difference in English and Math subscores, Bettinger et al., 2013).

### *Differences Between Good and Poor Comprehenders*

*Differences in MOCCA-College.* Data from the current study suggest that good and poor comprehenders in postsecondary institutions differ in a number of ways as reflected in their performance on MOCCA-College. First, poor comprehenders selected more PAR and ELA responses than did good comprehenders. Second, poor comprehenders had lower reading comprehension efficiency compared with good comprehenders. That is, on average, it took poor comprehenders longer to answer items correctly than it did good comprehenders—although poor comprehenders spent less time on the test overall. Third, good and poor comprehenders differed both in total test time and in the number of items completed. Good comprehenders spent more time on the test and answered more questions than did poor comprehenders. These results suggest that good and poor comprehenders may depend on different processing skills to make their selection on MOCCA-College. These data support the theoretical foundation used to develop MOCCA regarding *how* readers develop their situation models of reading (e.g., Kintsch, 2019) and extend previous research on how to identify this information with an efficient diagnostic assessment rather than through other more laborious methods (e.g., think aloud) that have been traditionally used to reach similar results (e.g., Magliano et al., 2011; McNamara, 2017).

#### *Differences in validity measures*

*ACT and SAT.* For all admissions data comparisons (i.e., ACT composite, ACT reading, ACT Math, SAT total, SAT reading, and SAT math), there were significant differences between good and poor comprehenders. These differences, in part, build preliminary validity evidence that MOCCA-College can distinguish between good and poor comprehenders. These findings extend the field by providing preliminary evidence of a new assessment that measures underlying skills and processes needed for reading comprehension (e.g., Biancarosa et al., 2019; Carlson et al., 2014; Davison et al., 2019) and, thus, provides informative information that could be used to make instructional decisions.

*NDRT.* A subset of participants provided responses to additional validation tasks by completing the NDRT. Small but significant differences were observed between comprehension skill groups on the NDRT total reading score. However, these findings should be interpreted with caution, given the small sample size, differences from the normative NDRT sample (Fishco, 2018), and the method used to identify good and poor comprehenders, which was based on scores from MOCCA-College itself. Future research may need to identify other criterion variables to better differentiate between comprehension skill groups.

### *Limitations*

Although the findings from the current study are promising, there are a number of caveats to these results. First, our study used a convenience sample mostly from three medium-to-large, regional, comprehensive universities. Although the sample was, at least from a demographic perspective, relatively representative of U.S. college students, it undoubtedly better reflects the population of students who matriculate into those colleges and universities than it does those who apply to them. As such, our assessment may better reflect the abilities and needs of students who are already enrolled in a postsecondary institution than those students applying to college. One goal of future studies involving MOCCA-College is to sample students from their senior year in high school to begin to further investigate college readiness.

In addition, our validation task sample reflects the fact that struggling comprehenders at the postsecondary level may be more reluctant to participate in reading studies. For instance, Harmon et al. (2005) found that adolescent struggling readers may be struggling with several issues such as seeing themselves as successful readers or learners, potentially creating anxiety or other social issues around seeking help to improve reading skills (Alvermann & Rush, 2004). Second, these findings do not reflect the abilities and comprehension differences of students enrolled (or co-enrolled) in 2-year institutions. Specifically, there is evidence to suggest that poor comprehenders are not monolithic and that these students may differ from their 4-year counterparts because of or despite their institution of choice (e.g., access, first-generation students, educational outcomes; Monaghan & Attewell, 2015; Pascarella et al., 1995, 2004; Somers et al., 2006).

Third, our NDRT subsample was not representative of the original NDRT normative sample (Fishco, 2018). Specifically, the NDRT normative sample included a mix of high school, 2-year college students, and 4-year college students, whereas our sample consisted entirely of 4-year college students. This difference helps explain why both good and poor comprehender groups had average scores on the NDRT that were higher than the average for the normative sample. Our NDRT subsample was also slightly more White, female, and younger than the NDRT normative sample. Additionally, our NDRT subsample was more White than our larger sample. However, we have confidence in using NDRT as a part of our cut scores because all three cut scores (i.e., ACT, SAT, and NDRT) showed convergence. Nonetheless, results based on NDRT must be interpreted with caution.

Fourth, as described earlier, one of the primary institutions has a test-optional or test-flexible policy. Consequently, analyses only reflect data from those students for whom admissions scores were available, thus limiting the generalizability of the inferences we can draw using the admissions test data. Given our inability to access data regarding disabilities or special education status, we are unable to further investigate why admission scores may not be reported. However, students with lower scores on admission tests may be less likely to report those scores (e.g., Robinson & Monks, 2005). We are unaware of any study examining test-optional or test-flexible policies specific to applicants with any disability or special education diagnoses. Nonetheless, although test-optional and test-flexible policies appear to increase application rates for historically underrepresented students and students from low socioeconomic status backgrounds (Epstein, 2009; Espenshade & Chung, 2011), they may not actually do much to increase student body diversity (Belasco et al., 2015).

Finally, we were unable to collect information about current accommodations or former special education status. Specifically, federal regulations and protections for students with special needs in college are different from those for students in K through Grade 12 settings. While students in K–12 settings may qualify for special services based on one of 13 Individuals with Disability Education Act Amendments of 1997 (IDEA, 1997) categories, this same category system does not apply to postsecondary education. However, postsecondary students may have a 504 plan. Nonetheless, as indicated above, it is impossible based on the current sample to draw conclusions related to individuals with learning disabilities.

### *Future Directions*

Although the findings reported here provide some preliminary evidence of the reliability of MOCCA-College, its utility in identifying meaningful differences in comprehender subgroups, and the validity of inferences regarding comprehension processes, these findings suggest the need for additional research. For example, future forms likely need to include a higher proportion of more difficult items, as measured by item statistics, readability, and genre (i.e., a higher proportion of expository items). Additionally, for many students, the assessment is too long to complete in a traditional class setting. Depending on how reliable shorter test forms prove to be, the number of items may be reduced to increase form completion. Alternatively, future versions of MOCCA-College could utilize a computer adaptive testing approach to reduce testing time and improve estimates of all three subscores (i.e., CCI, PAR, and ELA).

From a methodological perspective, our existing convenience sample draws heavily from students already enrolled in traditional 4-year institutions. Moving forward, we plan to recruit students enrolled in both community colleges and high schools. However, these students may represent somewhat different populations in terms of career paths, aptitude, and participant background. It is possible that MOCCA-College could capture those differences in college reading readiness and aid in interventions based on their cognitive processing needs. Similarly, future data collection will need to include mechanisms to access or request students' disability status (i.e., learning disabilities) and their indicated accommodations and modifications, as applicable.

Future validation work for MOCCA-College will also need to address the evolving nature of *test optional* and *test flexible* admissions and placement testing (Buckley, 2020). If admissions tests are not available in future years, we plan to request GPA as a criterion for validity analyses. From an analytic perspective, this would provide an opportunity to examine the extent to which the threshold scores identified in the current study generalize to subsequent samples and outcomes and allow for an evaluation of the feasibility and implications of alternative threshold scores. Likewise, additional validation measures (i.e., The Group Reading Assessment and Diagnostic Evaluation [GRADE], or the Gates-MacGinitie Reading Test) could be used to compare respective reader comprehension profiles and suggested interventions.

## Conclusion

The results of the current study suggest that MOCCA-College is a promising measure that does identify which postsecondary students struggle with reading comprehension, and our findings provide additional information as to *why* they struggle. As indicated, many college-bound students are not considered reading-ready by the benchmarks put forth by the ACT (2014). These results indicate that many postsecondary students struggle with comprehension because they may be failing to make the necessary causal connections while reading, and correlations found in this study between these data and other reading measures support a continued need to focus on the cognitive processes of reading comprehension to help improve students' comprehension skills. Specifically, MOCCA-College may help identify which postsecondary students need additional support in generating cognitive processes during reading comprehension.

### Acknowledgments

The authors thank the research assistants and project managers for their assistance in collecting the data presented in this article. They also thank the reviewers for their comments and suggestions toward improving the research in this article.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research presented in this manuscript was funded by research grant #R305A180417 from the Institute of Education Sciences (U.S. Department of Education) for "Multiple-choice Online Causal Comprehension Assessment for Postsecondary Students (MOCCA-College): Measuring Individual Differences in Reading Comprehension Ability of Struggling College Readers by Text Type."

### ORCID iDs

Ben Seipel https://orcid.org/0000-0002-0279-9480

Patrick C. Kennedy https://orcid.org/0000-0002-5525-3983

### Note

1. An ANOVA was used to examine whether examinees who were excluded from subsequent analyses (<10 responses on MOCCA-College and/or missing demographic data) differed from included examinees based on scores on either ACT Reading or SAT Reading admissions data. Raw scores were sample *z*-scored to allow for a pooled comparison across assessments. Admissions data were available for 152 of the 232 excluded students (66%) and 567 of 988 included students (57%). There were no significant differences between the groups, $F(1, 715) = 2.19$, $p > .05$, $\eta^2 < .01$.

### References

ACT. (2014). *The ACT® technical manual*. https://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf

ACT. (2016). *The condition of college and career readiness*. http://www.act.org/content/dam/act/unsecured/documents/CCCR_National_2016.pdf

ACT. (2021). *The condition of college and career readiness benchmarks*. https://www.act.org/content/act/en/college-and-career-readiness/benchmarks.html

Allen, J., & Radunzel, J. (2017). What are the ACT® college readiness benchmarks? *Reading*, *54*(75), 11–15. https://www.act.org/content/dam/act/unsecured/documents/pdfs/ACE-what-are-the-ACT-college-benchmarks.pdf

Alvermann, D. E., & Rush, L. S. (2004). Literacy intervention programs and the middle and high school levels. In T. L. Jetton & J. A. Dole (Eds.), *Adolescent literacy research and practice* (pp. 210–227). The Guilford Press.

Belasco, A. S., Rosinger, K. O., & Hearn, J. C. (2015). The test-optional movement at America's selective liberal arts colleges: A boon for equity or something else?. *Educational Evaluation and Policy Analysis*, *37*(2), 206–223. https://doi.org/10.3102/0162373714537350

Bergey, B. W., Deacon, S. H., & Parrila, R. K. (2017). Metacognitive reading and study strategies and academic achievement of university students with and without a history of reading difficulties. *Journal of Learning Disabilities*, *50*(1), 81–94. https://doi.org/10.1177/0022219415597020

Bettinger, E. P., Evans, B. J., & Pope, D. G. (2013). Improving college performance and retention the easy way: Unpacking the ACT exam. *American Economic Journal*, *5*(2), 26–52. https://doi.org/10.1257/pol.5.2.26

Biancarosa, G., Davison, M., Carlson, S. E., & Seipel, B. (2016, July). *Diagnosing the reading comprehension processes of poor comprehenders: Year 2 results of the multiple-choice online causal comprehension assessment* [Paper presentation]. Annual Meeting of the Society for the Scientific Study of Reading, Porto, Portugal.

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. https://doi.org/10.1177/0013164418763255

Biden, J. (2011, March 22). *Vice President Biden issues call to action to boost college graduation rates nationwide. Vice president releases college completion tool kit to help chart the path* [Press release]. https://www.whitehouse.gov/the-press-office/2011/03/22/vice-president-biden-issues-call-action-boost-college-graduation-rates-n

Boylan, H. R. (2009). Targeted intervention for developmental education students (T.I.D.E.S.). *Journal of Developmental Education*, 32(3), 14–18, 20, 22–23. https://files.eric.ed.gov/fulltext/EJ868669.pdf

Buckley, J. (2020). Standardized tests can serve as a neutral yardstick. *Education Next*, *20*(3), 67–68, 70, 72. https://www.educationnext.org/wp-content/uploads/2022/01/ednext_XX_3_forum.pdf

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*(5–6), 489–503. https://doi.org/10.1023/A:1008084120205

Cain, K., & Oakhill, J. V. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, *76*(4), 683–696. https://doi.org/10.1348/000709905X67610

Carlson, S. E., Seipel, B., & McMaster, K. (2014). Development of a new reading comprehension assessment: Identifying comprehension differences among readers. *Learning and Individual Differences*, *32*, 40–53. https://doi.org/10.1016/j.lindif.2014.03.003

College Board. (2012). *2012-13 educator's handbook for the SAT and SAT subject tests*. College Board SAT Program. https://vi.scribd.com/doc/133456894/Educators-Handbook-for-Sat-and-Sat-Subject-Tests

College Board. (2021). *Benchmarks*. College Board SAT Program. https://collegereadiness.collegeboard.org/about/scores/benchmarks

Colver, M., & Fry, T. (2016). Evidence to support peer tutoring programs at the undergraduate level. *Journal of College Reading and Learning*, *46*(1), 16–41. https://doi.org/10.1080/10790195.2015.1075446

Davis, O. S., Band, G., Pirinen, M., Haworth, C. M., Meaburn, E. L., Kovas, Y., Harlaar, N., Docherty, S. J., Hanscombe, K. B., Trzaskowski, M., Curtis, C. J., Strange, A., Freeman, C., Bellenguez, C., Su, Z., Pearson, R., Vukcevic, D., Langford, C., Deloukas, P., & Spencer, C. C. A. (2014). The correlation between reading and mathematics ability at age twelve has a substantial genetic component. *Nature Communications*, *5*(1), 1–6. https://doi.org/10.1038/ncomms5204

Davison, M. L., Biancarosa, G., Carlson, S. E., Seipel, B., & 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*, *43*(3), 169–181. https://doi.org/10.1177/1534508417728685

Davison, M. L., Biancarosa, G., Seipel, B., Carlson, S. E., Liu, B., & Kennedy, P. C. (2019). *Administration, interpretation, and technical manual 2019: Multiple-Choice Online Comprehension Assessment, MOCCA technical report MTR-2019* [Unpublished manuscript].

Davison, M. L., Seipel, B., Clinton, V., Carlson, S. E., & Kennedy, P. C. (2020). MOCCA college: An assessment of inferential narrative and expository comprehension. *Proceeding of International Conference on Higher Education Advances, Spain*, *6*, 417–425. https://doi.org/10.4995/HEAd20.2020.11081

Dorans, N. J. (1999). Correspondences between ACT™ and SAT® I scores. *ETS Research Report Series*, *1999*(1), i–18. https://doi.org/10.1002/j.2333-8504.1999.tb01800.x

Epstein, J. P. (2009). Behind the SAT-optional movement: Context and controversy. *Journal of College Admission*, 204, 8–19. https://eric.ed.gov/?id=EJ856358

Espenshade, T. J., & Chung, C. Y. (2011). Diversity outcomes of test-optional policies. In J. A. Soares (Ed.), *SAT wars: The case for test-optional admissions* (pp. 177–200). Teachers College Press.

Fike, D. S., & Fike, R. (2008). Predictors of first-year student retention in the community college. *Community College Review*, *36*, 68–88. https://doi.org/10.1177/0091552108320222

Fishco, V. V. (2018). *Technical manual for the Nelson-Denny reading test forms I and J*. Pro-Ed.

Goff, D. A., Pratt, C., & Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. *Reading and Writing*, *18*(7–9), 583–616. https://doi.org/10.1007/s11145-004-7109-0

Gorzycki, M., Howard, P., Allen, D., Desa, G., & Rosegard, E. (2016). An exploration of academic reading proficiency at the university level: A cross-sectional study of 848 undergraduates. *Literacy Research and Instruction*, *55*(2), 142–162. https://doi.org/10.1080/19388071.2015.1133738

Hanson, M. (2021, June 25). *College enrollment & student demographic statistics*. EducationData.org. https://educationdata.org/college-enrollment-statistics

Harmon, J. M., Keehn, S., & Kenney, M. S. (2005). Tutoring struggling adolescent readers: A program investigation. *Reading Research and Instruction*, *44*(2), 46–74. https://doi.org/10.1080/19388070409558426/

Hedges, L. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.2307/1164588

Henry, D. L., Nistor, N., & Baltes, B. (2014). Examining the relationship between math scores and English language proficiency. *Journal of Educational Research and Practice*, 4(1), 11–29. https://files.eric.ed.gov/fulltext/EJ1118452.pdf

Holschuh, J. P. (2019). College reading and studying: The complexity of academic literacy task demands. *Journal of Adolescent & Adult Literacy*, *62*(6), 599–604. https://doi.org/10.1002/jaal.876

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). John Wiley.

Individuals with Disability Education Act Amendments of 1997. (1997). https://www.congress.gov/105/plaws/publ17/PLAW-105publ17.pdf

Kincaid, J. P., Fishburne, L. R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease* *formula) for navy enlisted personnel* (Research Branch Report No. 8-75). Naval Air Station Memphis. https://apps.dtic.mil/sti/pdfs/ADA006655.pdf

Kintsch, W. (2019). Revisiting the construction–integration model of text comprehension and its implications for instruction. In D. E. Alvermann, N. J. Unrau, M. Sailors, & R. B. Ruddell (Eds.), *Theoretical models and processes of literacy* (pp. 178–203). Routledge. https://doi.org/10.4324/9781315110592-12

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Liu, B., Kennedy, P. C., 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*(4), 815–835. https://doi.org/10.1111/jedm.12238

Magliano, J. P., Millis, K. K., Levinstein, I., & Boonthum, C. (2011). Assessing comprehension during reading with the reading strategy assessment tool (RSAT). *Metacognition and Learning*, *6*(2), 131–154. https://doi.org/10.1007%2Fs11409-010-9064-2

Magliano, J. P., Millis, K. K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 107–136). Lawrence Erlbaum.

McMaster, K. L., Kendeou, P., Bresina, B., Slater, S., Wagner, K., White, M. J., Butterfuss, R., Kim, J., & Umana, C. (2019). Developing an interactive software application to support young children's inference-making. *L1-Education Studies in Languages and Literature*, *19*, 1–30. https://doi.org/10.17239/L1ESLL-2019.19.04.04

McMaster, K. L., van den Broek, P., Espin, C. A., White, M. J., Rapp, D. N., Kendeou, P., Bohn-Getler, C. M., & Carlson, S. E. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences*, *22*, 100–111. https://doi.org/10.1016/j.lindif.2011.11.017

McNamara, D. S. (2017). Self-explanation and reading strategy training (SERT) improves low-knowledge students' science course performance. *Discourse Processes*, *54*(7), 479–492. https://doi.org/10.1080/0163853X.2015.1101328

McNamara, D. S., & Kendeou, P. (2011). Translating advances in reading comprehension research to educational practice. *International Electronic Journal of Elementary Education*, 4(1), 33–46. https://files.eric.ed.gov/fulltext/EJ1068606.pdf

Monaghan, D. B., & Attewell, P. (2015). The community college route to the bachelor's degree. *Educational Evaluation and Policy Analysis*, *37*(1), 70–91. https://doi.org/10.3102/0162373714521865

Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, *9*(4), 329–389. https://doi.org/10.1207/s1532690xci0904_2

National Center on Intensive Intervention. (2020). *Academic screening tools chart rating rubics*. https://intensiveintervention.org/sites/default/files/NCII_AcademicScreening_RatingRubric_2020-06-30.pdf

Pascarella, E. T., Bohr, L., Nora, A., & Terenzini, P. (1995). Cognitive effects of 2-year and 4-year colleges: New evidence. *Educational Evaluation and Policy Analysis*, *17*(1), 83–96. https://doi.org/10.2307/1164271

Pascarella, E. T., Pierson, C. T., Wolniak, G. C., & Terenzini, P. T. (2004). First-generation college students: Additional evidence on college experiences and outcomes. *The Journal of Higher Education*, *75*(3), 249–284. https://doi.org/10.1080/00221546.2004.11772256

Poole, A. (2019). Reading strategies and academic success: The case of first-semester college males. *Journal of College Student Retention*, *21*(1), 2–20. https://doi.org/10.1177/1521025116685094

Qualtrics. (2019). *Qualtrics* [Computer software]. Qualtrics.

Robinson, M., & Monks, J. (2005). Making SAT scores optional in selective college admissions: A case study. *Economics of Education Review*, *24*, 393–405. https://doi.org/10.1016/j.econedurev.2004.06.006

Ronimus, M., Eklund, K., Pesu, L., & Lyytinen, H. (2019). Supporting struggling readers with digital game-based learning. *Educational Technology Research and Development*, *67*(3), 639–663. https://doi.org/10.1007/s11423-019-09658-3

Sadler, P. M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *Journal of Research in Science Teaching*, *35*, 265–296. https://doi.org/10.1002/(SICI)1098-2736(199803)35:3%3C265::AID-TEA3%3E3.0.CO;2-P

Scott-Clayton, J., & Rodriguez, O. (2014). Development, discouragement, or diversion? New evidence on the effects of college remediation policy. *Education Finance and Policy*, *10*, 4–45. https://doi.org/10.1162/EDFP_a_00150

Somers, P., Haines, K., Keene, B., Bauer, J., Pfeiffer, M., McCluskey, J., Settle, J., & Sparks, B. (2006). Towards a theory of choice for community college students. *Community College Journal of Research and Practice*, *30*(1), 53–67. https://doi.org/10.1080/10668920500248886

Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist*, *47*(4), 522–532. https://doi.org/10.1037/0003-066X.47.4.522

Tinto, V. (2012). *Completing college: Rethinking institutional action*. University of Chicago Press.

U.S. Census Bureau. (2018, December 11). *Classroom diversity on the rise*. https://www.census.gov/library/visualizations/2018/comm/classroom-diversity.html

Varma, S. (2006). *Preliminary item statistics using point-biserial correlation and p-values*. Educational Data Systems. https://eddata.com/wp-content/uploads/2015/11/EDS_Point_Biserial.pdf

Weathington, B. L., Cunningham, C. J. L., & Pittenger, D. J. (2012). *Understanding business research*. John Wiley.