

# The Efficacy of an Online Cognitive Assessment Tool for Enhancing and Improving Student Academic Outcomes

Lindsay Shaw, Janet MacIsaac, and Jill Singleton-Jackson  
*Department of Psychology, University of Windsor*

## Abstract

With technology at the fingertips of most undergraduate students, it has been difficult for instructors to fully engage students in the classroom, which has resulted in the creation of several innovative online cognitive assessment tools. These tools often integrate several cognitive learning strategies within an assessment, with the goal of actually enhancing learning, as opposed to just measuring it. In the current study, students' level of engagement and test performance using a recently developed online application were compared to their final multiple-choice paper-and-pencil exam mark to determine the efficacy of the new online application in achieving improved learning outcomes. Results indicated that students had high test scores using the online tool despite their limited engagement in the cognitive learning features, calling into question the online cognitive assessment tool's facilitation of long-term learning. Implications and recommendations for future online cognitive assessment application implementation in educational environments are discussed.

*Keywords:* online assessment, student engagement with online tools, online learning

Shaw, L., MacIsaac, J., & Singleton-Jackson, J. (2019). The efficacy of an online cognitive assessment tool for enhancing and improving student academic outcomes. *Online Learning, 23*(2), 124-144. doi:10.24059/olj.v23i2.1490

## The Efficacy of an Online Cognitive Assessment Tool for Enhancing and Improving Student Academic Outcomes

A new challenge educators face in university classrooms is student distractibility due to the influence of technology and social media. Nevertheless, students are most engaged during the assessment component of education, which provides a unique opportunity for educators to enhance learning outcomes, and not just assess student knowledge. Traditionally, to assess student knowledge, we examine a student's ability to recall information (e.g., short-answer questions, essays) or recognize information (e.g., multiple-choice questions). The latter creates an issue with the assessment process because if we are actually interested in measuring students' understanding of a concept, recognition-based assessment methods make it difficult to accurately assess their knowledge. In other words, are we interested in a student's understanding and mastery of content or their ability to recognize a correct response? In fact, students have been identified as "passive actors" rather than "active learners" when it comes to such assessment practices (Chappuis, Stiggins, Arter, & Chappuis, 2004). Thus, there may be a clear disconnect between the goals of assessment and what is actually occurring.

## Factors Impacting the Assessment Process

Several student characteristics can impact the assessment process. First, students may vary in their academic orientations, which are broadly distinguished by learning-oriented (LO) and grade-oriented (GO) behaviors. A student that is high in LO places greater importance on actually learning the material, whereas a student high in GO places greater importance on receiving a higher grade. Many studies measure LO and GO as dichotomies with the expectation of students being high in one and low in the other (Beck, Rorrer-Woody, & Pierce, 1991; Frymier & Weser, 2001; Lawrence & Frymier, 2002; Williams & Frymier, 2007). However, researchers have found that it is possible to score high in both and low in both. Therefore, instead of having two distinct groups (high LO/low GO, high GO/low LO), there are four (high LO/low GO, high GO/low LO, high GO/high LO, low GO/low LO). The latter group, low GO/low LO, is very rare in an academic setting, as these individuals have little to no motivation for learning and are more likely in school for external reasons, such as parental expectation or socialization practices inculcating the importance of higher education to ensure success (Eison, Pollio, & Milton, 1986; Roedel, Schraw, & Plake, 1994). A student that is high in both LO and GO has the desire to learn with an expectation of receiving a high grade as a result of their engagement in the learning process (Marsden, Carroll, & Neill, 2005). Ultimately, less is known about the distinctive academic characteristics of these two groups, which has led researchers to explore students within the two dichotomies as opposed to the four categorizations.

Two other student characteristics can also impact the assessment process: academic entitlement (AE) and test anxiety (TA). An academically entitled student is an individual who demands higher grades for reasons independent of their performance (Kopp, Zinn, Finney, & Jurich, 2011). For instance, a student exhibiting high levels of AE may demand a high grade simply because they attend every class or do every reading. Of interest, students higher in AE have reported higher academic demands from their parental figures, which in turn creates more anxiety about grades. This shifts their focus away from learning and mastering the course content and toward greater grade achievement (Greenberger, Lessard, Chen, & Farruggia, 2008). This suggests that students higher in AE may experience more anxiety about their grades, which may lead them to adapt more of a GO than a long-term learning approach. Although the research on AE has been fairly recent, researchers have found several characteristics related to AE populations: higher levels of work avoidance, frustration/negative attitudes, and less motivation and responsibility for actions (Cain, Romanelli, & Smith, 2012).

Students with high GO tend to report higher levels of TA (Eison et al., 1986), whereas those with high AE tend to have more anxiety about their grades (Greenberger et al., 2008). This is particularly important in terms of assessment, as those with high TA have more trouble encoding and storing information (Everson, Smodlaka, & Tobias, 1995; Thomas, Cassady, & Heller, 2017), and are less efficient in cue-utilization strategies (Cassady & Johnson, 2002). Researchers have suggested that these poorly implemented learning strategies are a result of poor studying habits, such as procrastination (Kalechstein, Hocevar, Zimmer, & Kalechstein, 1989) or avoidant coping strategies, such as mental disengagement (Stoeber, 2004; Thomas, Cassady, & Heller, 2017; Zeidner & Matthews, 2005). However, students experiencing high TA may know the material just as well as their less anxious peers but have trouble utilizing the appropriate cognitive learning strategies to retrieve that information during assessment. This anxiety can be especially heightened in a testing environment that has time pressures (Plass & Hill, 1986), which is a characteristic of most assessments in higher education. Consequently, students with high TA often perform worse

on assessments than their peers with lower TA (Decaro et al., 2011; Thomas, Cassady, & Heller, 2017; Zeidner & Matthews, 2005) and attribute their failure to external sources (Cassady, 2004). This cycle of external attribution continues into future assessments and creates a negative perception of assessment, in turn creating a cycle of avoidant behaviors. If students can control their anxiety during assessments, their performance may increase, along with their perceptions of the assessment process.

### **Engagement and Performance**

More positive attitudes about assessment are important, as negative perceptions are rooted in students' understanding of learning. As such, students have unique perspectives on their preference for assessment and how they approach taking tests (Struyven, Dochy, & Janssens, 2005). For instance, multiple-choice pen-and-paper (or MCPP) tests have had a long history of students' utilizing a surface approach to learning (SAL), triggering more positive attitudes toward that assessment type in comparison to open-ended/essay formats (Furnham, Batey, & Martin, 2011; Scouller, 1998; Scouller & Prosser, 1994). This finding was particularly relevant for students with higher TA, suggesting that the underlying reason for this preference was not the increased learning associated with MCPP practices but rather that students found it easier to prepare for and take MCPP tests (Traub & McRury, 1990; Birenbaum & Feldman, 1998).

Since SAL contradicts the overall goal of assessments, various other assessment types have been implemented in postsecondary institutions, such as peer assessments, self-assessments, and online cognitive assessment tools (OCATs), such as MyPsychLab, all of which encompass the same goal of learning but in different forms. OCATs have benefits and drawbacks for both instructors and students. For instructors, the use of an OCAT allows easy distribution and marking of tests, which is especially important for large introductory courses. Yet becoming familiar with these online platforms can be challenging and result in a time-consuming learning curve. This latter drawback can also be applied to students who have never used an OCAT, which can heighten the stress and anxiety of tests even more (Özden, Ertürk, & Sanli, 2004). However, there is evidence to support the idea that heightened anxiety has decreased as online activities have become everyday practice (Dermo, 2009; Walker, Topping, & Rodrigues, 2008). OCATs can also be convenient because they can be used at a place and time suitable for the student. This adaptability can also contribute to reductions in anxiety. Additionally, the OCATs can provide quick and comprehensive feedback to students, allowing students to self-regulate their knowledge of tested content (Miller, 2009). The benefits of such assessments must outweigh the drawbacks, as students have reported positive attitudes toward the introduction of online assessments within their educational environments (Miller, 2009; Smith & Caruso, 2010).

With this in mind, there is a need for assessment tools that foster greater long-term learning for students. One newly developed OCAT has been created in an attempt to fill this gap in higher education (Pare & Joordens, 2009). The OCAT being assessed in the present study uses several different cognitive learning strategies to optimize the learning process. What makes this tool interesting is the approach it takes to enhancing the multiple-choice format. For the OCAT being assessed, the multiple-choice format begins with free recall and ends with immediate feedback, embedding retrieval cues and second opportunities to answer questions for fewer marks. Figure 1 displays this pattern for each multiple-choice question that students encounter.

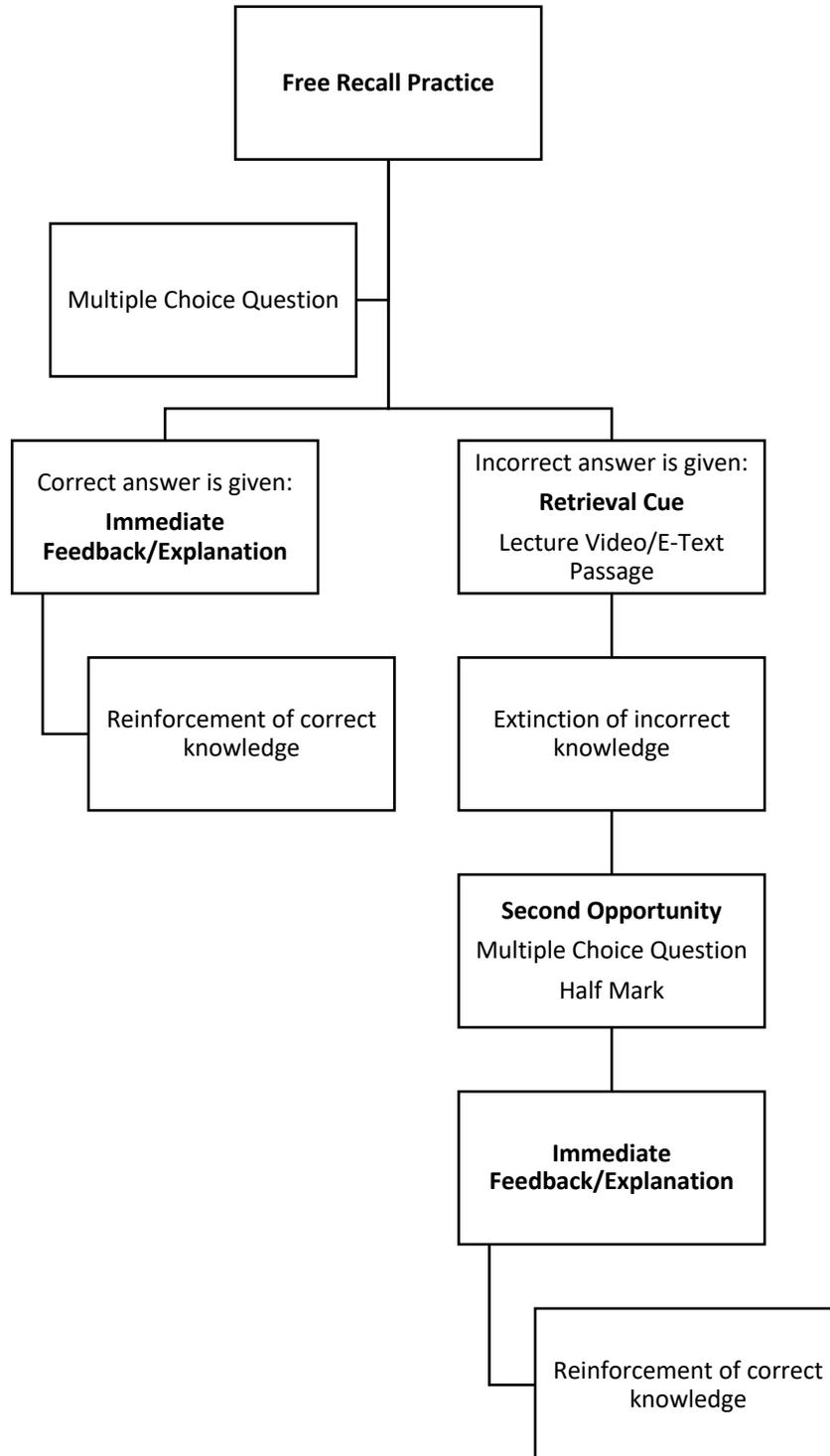


Figure 1. OCAT process with cognitive learning strategies.

**Retrieval to free recall.** These cognitive learning strategies enhance information retention and can strengthen long-term learning through testing, which is referred to as the *testing effect* (Carpenter & DeLosh, 2006). The strength of the testing effect is dependent on the type of retrieval and the underlying mechanisms required for that retrieval. Retrieval is the process of accessing knowledge, often from our stored memory with the help of environmental retrieval cues (Karpicke, 2012). Retrieval is essential for our learning because it not only helps us access knowledge, but it also helps enhance knowledge, making it easier to retrieve in the future (Roediger & Karpicke, 2006; Karpicke & Roediger, 2007, 2008; Karpicke & Zaromb, 2010). Recall tests (e.g., short answer, essay) strengthen long-term learning because the retrieval process is needed for success in recall but not for success in recognition (Carpenter & DeLosh, 2006). As students continue to practice recalling information, the retrieval routes that transfer the knowledge are strengthened (McDaniel & Masson, 1985). During the OCAT assessment, the first feature of each multiple-choice question requires students to engage in free recall, where they can type as much information as they want in the dialogue box. This acts as a shortened essay format where students can experience deep approach learning (DAL) and prime associations. The more they engage in the free recall feature, the more their retrieval routes are being used and hopefully strengthened.

**Unsuccessful retrieval attempts and immediate feedback.** If the retrieval route is already strengthened from previous practice, successful retrieval attempts (i.e., getting the correct answer) are more likely to be made. Yet unsuccessful retrieval attempts are less understood, specifically in how they benefit learning. Often, students write MCPP examinations, receive their marks weeks later, and rarely identify which items they got right and which they got wrong. This increases the likelihood of students learning erroneous concepts, impacting their performance on subsequent examinations (Marsh, Roediger, Bjork, & Bjork, 2007). However, unsuccessful retrieval can be counteracted if immediate feedback is provided (Kornell et al., 2009). In the OCAT assessment, students are given immediate feedback (i.e., explanations) for every question regardless of successful or unsuccessful attempts, which ultimately reinforces the correct concepts.

**Second chances and retrieval cues.** Retrieval is often cue dependent (Tulving, 1974). Students might fail to recall the information because of ineffective retrieval cues. These retrieval cues can come from the external or internal environment. If the cue helps the student further understand the concept, they can use that cue for studying practices and future examinations. In the OCAT assessment, an incorrect response to a multiple-choice question is followed by a retrieval cue (i.e., a hint) in the form of a brief video lecture or e-text passage. Students then have a second attempt to answer the same question for half the mark. Additional benefits of including multiple attempts within examinations include eliminating student anxiety and emphasizing the importance of student development, not just academic outcomes (Baleni, 2015).

## **OCAT Research**

The theoretical foundation of the OCAT being assessed has been well established, as the aforementioned cognitive learning strategies are shown to increase learning in various educational contexts. Yet no published research exists on this assessment tool despite its use in postsecondary institutions. However, Holbrook, Dupont, Power, and Joordens (2015) did conduct a pilot project that compared the OCAT to an online learning management system called LEARN using a multiple-choice format experiment, with students preferring the OCAT over LEARN—particularly enjoying the immediate feedback feature. LEARN, used at the University of Waterloo, is similar to other online learning management systems used in university settings, such as Blackboard or Moodle. Although this pilot study provides some insight into students' attitudes and

test performance with the OCAT, there are several gaps that still need further exploring. For instance, it is important to understand how students engage in the cognitive learning features of the OCAT, as this should be, in theory, an important indicator of their performance on the assessments.

## Present Research

The current study was initially carried out as part of a larger research project for a master's thesis.<sup>1</sup> However, the focus of this article is on one aspect of that project. We are reporting on our findings about students' level of engagement with the cognitive learning features, specifically in students with varying LO and GO, levels of AE, and TA. The research question we report on is this: Will students' combination of LO and GO, AE, and TA predict their level of engagement during each cognitive learning strategy (i.e., free recall, second attempt, immediate feedback, retrieval cue)?

**Hypothesis.** Student engagement was explored within each cognitive learning strategy. We hypothesized that students with higher levels of GO, TA, and AE would report lower levels of engagement in free recall and immediate feedback and higher levels of engagement in the second attempts feature and retrieval cues. The opposite was expected in students with higher LO.

## Methods

All students were recruited from the university's research pool, and they were treated in accordance with TCPS-2 ethical guidelines. The study was approved by the institutional research ethics board. Both introduction to psychology courses were taught by the same instructor, using the same content and the same evaluations, ensuring consistency across the sample.

## Measures

Participants completed two sets of online surveys (a presurvey and postsurvey) and were given a unique ID code, which was used to match participants over the survey sessions and semesters. These ID codes ensured that students would remain anonymous to the instructor in the course. Participants also provided additional demographic information, specifically their gender, age, ethnicity, program, and year of study. Performance was measured using students' final grades on the MCPP exam, their final grades in the course, and their final OCAT midterm grades. The lead researcher had access to the grades because of her role as the OCAT administrator and graduate assistant for the course but had no contact with the students in the course.

**LOGO-II.** Learning and grade orientations were assessed using the LOGO-II scale developed by Eison, Pollio, and Milton (1983). This scale had two sections, the first with 16 items regarding specific attitudes (e.g., "I dislike extra assignments that are not graded") on a 5-point Likert-type scale (1 = *strongly disagree* to 5 = *strongly agree*). The second section had 16 items regarding behaviors (e.g., "I cut classes when confident that lecture material will not be on the exam") on a 5-point Likert-type scale (1 = *never* to 5 = *always*). Recommended cutoffs from previous studies were then used to categorize students in their high/low dichotomies. The reliability of this 32-item scale ( $\alpha = .72$ ) was consistent with prior reliability ( $\alpha = .70$ ) analyses (Levine, 2003; Purcell, 2010).

---

<sup>1</sup> The interested reader can contact the researchers for a copy of the original thesis.

**Cognitive Test Anxiety scale.** The Cognitive Test Anxiety scale (CTA), which has 27 items, was used to measure students' TA (Cassady & Johnson, 2002). An example item was "I lose sleep over worrying about examinations," which was measured on a 4-point Likert-type scale (1 = *not at all typical of me* to 4 = *very typical of me*). Scores ranged from 27 to 108, with higher scores indicating higher TA. A cutoff point of 69 was used to dichotomize low and high levels. The reliability of this scale ( $\alpha = .83$ ) was lower than prior reliability ( $\alpha = .91-.93$ ) analyses (Cassady, 2004; Furlan, Cassady, & Perez, 2009).

**Academic Entitlement Questionnaire.** AE was assessed using the 8-item Academic Entitlement Questionnaire (AEQ), which is a smaller, theoretically constructed measure from the original that had 26 items (Kopp et al., 2011). It is measured using a 5-point Likert-type scale (1 = *strongly disagree* to 5 = *strongly agree*). As such, students' scores can range from 8 to 40, with a cutoff point of 24. Scoring above that cutoff point would result in high AE; scoring below that cutoff point would result in low AE. The reliability of this 8-item scale was higher ( $\alpha = .83$ ) than that found in previous studies ( $\alpha = .80$ ; Kopp et al., 2011).

**Engagement scale.** An engagement scale was created to assess engagement with the OCAT's built-in cognitive learning strategies. Three items were adapted from the National Survey of Student Engagement (NSSE) scale, which measures the quality of engagement of students across universities in the United States and Canada, producing an overall internal reliability of .8 (Tendhar, Culver, & Burge, 2013). All other items were developed based on the theoretical understanding of the engagement construct, which has been defined as the amount of time students devote to a desired outcome (Kuh, 2001) as well as their level of collaboration and communication with staff and peers (Coates, 2007). Therefore, the scale contained eight items, four of which measured how often students engaged in each cognitive learning strategy and the other four measuring their level of engagement with their peers and professor, with an example item being "During the current OCAT midterm, generally how often did you read the feedback given after each multiple-choice question?" The items were measured on a 5-point Likert-type scale (from 1 = *never* to 5 = *all the time*). The combined items produced a low reliability score ( $\alpha = .64$ ). Therefore, the items were used individually for all subsequent analyses.

## Procedure

The study ran from October 2016 until May 2017, covering two academic semesters. Depending on students' program and interest in the introductory psychology course, students could have been enrolled in both semesters or just one. For example, a student could enroll in just the fall semester, meaning they would complete the study protocol during that semester. Another student could enroll in just the winter semester, meaning they would complete the following protocol in that semester. Finally, participants could have signed up for both semesters, in which case they would complete the presurvey in the first semester and then complete the postsurvey in the second semester. The present report is focused solely on the engagement with the OCAT's built-in cognitive learning strategies. As such, the study flow allowed participants to first gain exposure to the platform while simultaneously measuring relevant student characteristics (e.g., LO, GO, TA, and AE) and student performance (OCAT midterm results, final MCPP grade, and final course grade) to compare against the postsurvey responses for engagement with the cognitive learning strategies.

Thus, participants in the study completed the following protocol: Each semester, eligible participants completed the presurvey containing the questionnaires focused on learning approach

(LOGO-II), test anxiety (CTA), and academic entitlement (AEQ). The presurvey was available for the duration of the study to accommodate participants across the semesters if they hadn't already completed the presurvey (e.g., a student signs up for just the second semester course). Then, during each semester, participants would complete three midterms using the OCAT instead of the traditional MCPP midterms used in the course, followed by a final MCPP exam. Prior to the first OCAT midterm in each semester, students were provided with instructions that would assist them in using the OCAT during the midterm assessment process.<sup>2</sup> The instructor also gave detailed instructions on what the assessment was, the importance of the assessment, and how it was graded, during a lecture at the beginning of each semester. Finally, participants completed the postsurvey in either December 2016 or May 2017 depending on whether they enrolled in one semester or both semesters. During the postsurvey, participants were asked about their engagement with the OCAT's learning features over the course of the semester. For both semesters, participants' responses on the presurvey and postsurvey, their OCAT midterm grades, their final MCPP grade, and their final course grade were used to examine the research hypothesis.

### Sample Description

In total, 410 participants enrolled in a traditional, on-campus first-year introduction to psychology course at medium-sized university campus in southern Ontario completed the presurvey. Unfortunately, the study had drastic attrition, as only 155 students completed the postsurvey, which qualifies our final analysis. Nevertheless, this smaller sample mirrored much of the presurvey sample, as the majority identified as Caucasian/European (69.1%) women (77.1%) in their first year of study ( $M_{Age} = 19.46$ ,  $SD_{Age} = 2.12$ ) in the arts, humanities, and social sciences departments (59.2%). Yet many of the groupings for the student orientations changed, with the majority having high LO (62.6%) and high GO (54.8%). Given the predictive and exploratory nature of the research questions, all results pertaining to those questions will use this postsurvey sample. Although, in hindsight, there could have been strategies put in place to eliminate the amount of attrition. As such, there were various consequences to this procedure, which are discussed later in the limitations section. See Table 1 and Table 2 for the postsurvey sample description. See Table 3 for student learning characteristic means and standard deviations.

---

<sup>2</sup> Please contact the researchers for a copy of the instructions.

Table 1  
*Postsurvey Sample Characteristics*

	Frequency	Percentage of Sample
<b>Ethnic Background*</b>		
Caucasian/European	103	69.1
Middle Eastern	16	10.7
Asian	9	6.0
African American	9	6.0
East Indian	6	4.0
Mixed	5	3.4
Hispanic	1	0.7
<b>Discipline**</b>		
Arts, humanities, social sciences	90	59.2
Science	31	20.4
Business	14	9.2
Human kinetics	7	4.6
Double major	7	4.6
Nursing	3	2.0

Note. \*( $n = 149$ ) as six participants chose not to specify their ethnicity. \*\*( $n = 152$ ) as three participants chose not to specify their academic major.

Table 2  
*Student Orientations (Postsurvey)*

Student Orientation	Designation	Frequency	Percentage of Sample
AE*	High	25	16.7
	Low	125	83.3
GO	High	85	54.8
	Low	70	45.2
LO	High	97	62.6
	Low	58	37.4
TA**	High	53	37.1
	Low	90	62.9

Note. \*( $n = 150$ ) as five participants chose not to complete the survey. \*\* ( $n = 143$ ) as 12 participants chose not to complete the survey.

Table 3  
*Descriptive Statistics for Student Characteristics*

Student Characteristics	$n$	Minimum	Maximum	Mean ( $SD$ )
TA	143	38	86	64.50 (10.64)
AE	150	8	34	18.49 (5.37)
GO	151	24	72	46.15 (7.60)
LO	151	28	72	46.66 (6.33)

## Results

### Overview

All data sources were analyzed using SPSS software. Assumptions were checked for each test, and unless otherwise stated, these assumptions were met. Outliers and influential observations were assessed using standardized residual scores ( $\pm 2.5$ ), leverage scores, and Cooks distance ( $< 1$ ). Our findings are presented in three parts. First, we report on participants' self-reported engagement with the OCAT. Second, a series of multiple regressions were carried out to examine how participants' TA, AE, LO, and GO predicted engagement with the OCAT cognitive strategies. Finally, we report on the difference between the OCAT grades and final MCPP exam grade.

### Research Question

**Engagement.** The first engagement construct measured the amount of time students devoted to each cognitive learning strategy on a 5-point Likert-type scale. Students “only sometimes” engaged in the cognitive learning features, specifically the immediate feedback ( $M = 3.28$ ,  $SD = 1.15$ ), retrieval cues ( $M = 2.95$ ,  $SD = 1.35$ ), and the free recall ( $M = 2.59$ ,  $SD = 1.26$ ). The limited engagement in the latter cognitive learning feature aligns closely with students' attitudes, as only 5.8% preferred this feature over the others. Yet the highest engagement for students was the second opportunity feature, where they used it almost every time ( $M = 3.98$ ,  $SD = 1.22$ ). Similarly, this engagement aligns closely with student attitudes, as the second opportunity feature was preferred most by the students.

Additional questions also explored students' amount of collaboration and communication with staff and peers. Before the assessment, students indicated that they “almost never” provided help to their peers ( $M = 2.43$ ,  $SD = 1.13$ ), sought help from their peers ( $M = 2.14$ ,  $SD = 1.12$ ), or sought help from the professor about the OCAT ( $M = 1.85$ ,  $SD = 1.03$ ). Interestingly, this lack of communication and collaboration extended into the assessment process, with students indicating that they “almost never” worked with their peers during the OCAT assessment ( $M = 2.08$ ,  $SD = 1.36$ ). However, it is important to note that these results came from students' self-reports, and students are instructed to not collaborate on tests. Therefore, the validity of the assessment of the collaborative aspect is questionable and may not resemble how the students actually worked during the assessment process.

**Predicting engagement in immediate feedback.** Four outliers were identified, but running the analysis with and without them did not make any significant changes to the model or  $R^2$  value. This, paired with the smaller sample size, resulted in all outliers being kept in the analysis.

The model, which contained the four predictor variables, was statistically significant,  $F(4, 137) = 3.34$ ,  $p < 0.05$ ,  $R^2 = 0.09$ ,  $R^2_{\text{Adjusted}} = 0.06$ , accounting for only 9% of the variance in the model. However, LO ( $M = 46.65$ ,  $SD = 6.38$ ) was the only statistically significant predictor,  $b = 0.04$  ( $SE = 0.02$ ),  $\beta = .21$ , 95% CI [0.01, 0.07],  $t(141) = 2.53$ ,  $p < .05$ . This indicates that an increase in LO would result in heightened engagement in the immediate feedback feature. In contrast, AE, GO, and TA did not predict engagement in the immediate feedback feature. This provides partial support of the initial hypothesis of LO students having higher engagement in the immediate feedback feature while students with high TA, GO, and AE would have less engagement in the feature.

**Predicting engagement in retrieval cues.** Four outliers were identified, but running the analysis with and without them did not make any significant changes to the model or  $R^2$  value. Therefore, the outliers were kept in the analysis.

The model, which contained the four predictor variables, was statistically significant  $F(4, 137) = 5.86, p < 0.001, R^2 = 0.15, R^2_{Adjusted} = 0.12$ , though only 15% of variance was accounted for. There were two statistically significant predictors (see Table 4 for results). The two predictor variables had differing directional relationships. An increase in LO led to an increase in the retrieval cue engagement, whereas an increase in GO led to a decrease in the retrieval cue engagement. However, AE did not predict engagement in the retrieval cue feature. Though it is important to note that TA was approaching statistical significance. It was initially hypothesized that students with high TA, GO, and AE would have more engagement in the retrieval cue feature, as it contributes to a higher grade. Additionally, it was hypothesized that students with high LO would engage less in the retrieval cue. However, the opposite occurred for both LO and GO students.

Table 4  
*Regression Model for Engagement in Retrieval Cues*

Predictors	<i>b</i>	<i>SE</i>	$\beta$	<i>t</i>	<i>p</i>	<i>R</i>	$R^2$	Adjusted $R^2$
AE	.00	.02	.01	.02	.98	.38	.15	.12
GO	-.06	.02	-.33	-3.62	< .01			
LO	.04	.02	.20	2.39	< .05			
TA	.02	.01	.15	1.87	.06			

**Predicting engagement in second chance feature.** In assessing the assumptions, all were satisfied. The model, which contained the four predictor variables, was not statistically significant,  $F(4, 136) = 1.33, p > 0.05$ . With that, students' AE, GO, LO, and TA did not predict their level of engagement in the second chance feature not supporting the initial hypothesis. It was hypothesized that students with higher AE, GO, and TA would engage more in the second attempt feature because it helps contribute to a higher grade. The opposite was expected for students with high LO, as they would have been more academically prepared and would not need to use this feature as often as their peers.

**Predicting engagement in free recall.** A total of five outliers were identified. Running the analysis with and without them did not make any significant changes to the model or  $R^2$  value. Therefore, all outliers were kept in the analysis.

The model, which contained the four predictor variables, was statistically significant  $F(4, 137) = 6.80, p < 0.001, R^2 = 0.17, R^2_{Adjusted} = 0.14$ , though it only accounted for 17% of the variance. Additionally, all four predictors were statistically significant (see Table 5 for results). The latter two predictor variables had a negative relationship, indicating that an increase in GO and TA led to a decrease in engagement in the free recall feature. In contrast, LO and AE had a positive relationship, indicating that an increase in LO and AE led to an increase in engagement in the free recall feature. This partially supported the initial hypothesis. It was initially hypothesized that students with high LO would have more engagement in free recall and that GO, TA, and AE would engage less in this feature.

Table 5  
*Regression Model for Engagement in Free Recall*

Predictors	<i>b</i>	<i>SE</i>	$\beta$	<i>t</i>	<i>p</i>	<i>R</i>	<i>R</i> <sup>2</sup>	Adjusted <i>R</i> <sup>2</sup>
AE	.05	.02	.20	2.29	<.05	.41	.17	.14
GO	-.03	.02	-.19	-2.11	<.05			
LO	.07	.02	.33	4.12	<.01			
TA	-.02	.01	-.18	-2.20	<.05			

**Performance.** Although the sample was typical of most university classrooms, the final grades were higher ( $M = 79.79$ ,  $SD = 9.30$ ), which was likely because of extremely high OCAT grades ( $M = 90.60$ ,  $SD = 8.91$ ) balanced with average MCPP final exam scores ( $M = 65.40$ ,  $SD = 13.00$ ). A paired samples *t*-test was used to identify whether the difference between the final MCPP grades and final OCAT grades was statistically significant. A boxplot inspection revealed four outliers and two influential observations. The outliers and influential observations were present on the OCAT grades variable, except for one outlier on the final MCPP variable. The data were analyzed both with and without the outliers and influential observations. Although both analyses produce similar results, the removal of the outliers and influential observations fixed some of the normality issues. Therefore, the outliers and influential observations were Winsorized to meet the second highest data point. A positive skew was still present through visual and statistical inspections. Since the test is typically robust to normality violations and the integrity of the data wanted to be kept, the data was not transformed. As expected, the difference between the two assessment practices was statistically significant,  $t(121) = 21.51$ ,  $p < 0.001$ , 95% CI [23.06, 27.73], and had a large-sized effect,  $d = .88$ . Thus, students performed significantly better on the OCAT assessments ( $M = 90.83$ ,  $SD = 7.88$ ) than on the MCPP assessment ( $M = 65.44$ ,  $SD = 12.89$ ).

## Discussion

Technology, often literally at the fingertips of most young adults, can help students decrease their reliance on memorization and equip them with the necessary cognitive learning strategies to succeed in the classroom. As such, the OCAT, like most new innovative assessments, was specifically developed to meet the growing needs of this incoming generation by steering them away from the passive nature of traditional MCPP assessments while moving them toward being more active, responsible learners. Although the OCAT was built on a strong theoretical background that supports learning, it was important to identify whether different academically oriented students actually engaged in the assessment and whether they actually learned from the assessment. Explorations into student characteristics was also important to pursue in the hope that it would, in part, fill in the various gaps within the current literature.

## Engagement

Engagement was low for the students, in their communication and collaboration with the professor and peers. However, students were instructed not to work with one another during the assessment, which potentially explains why peer collaboration during assessment was low. However, this study did rely heavily on self-reports, which doesn't rule out peer collaboration from taking place. Additionally, the ease of navigating through the OCAT may have reduced students' need to seek advice from their professor and peers. The more striking findings were students' limited engagement in the four cognitive learning features.

**Always prompted: Free recall and immediate feedback.** The OCAT is structured such that each student is given the opportunity to engage in free recall and immediate feedback for every multiple-choice question. Thus, if students were utilizing the cognitive learning strategies optimally, then they would have identified that they used these two features “all of the time.” Instead, the majority “only sometimes” used these features. This suggests that students were, at times, quickly skipping the free recall, which was their first prompt, and quickly skipping the immediate feedback, which was their last prompt. Skipping the free recall feature was likely a function of three things. First, the free recall feature was the most strenuous task for the students during the assessment. Perhaps the students did not want to put in the extra effort to engage in it. Second, they may not have understood the importance of their engagement on the learning process. And third, they may not have understood what to do with the box in the first place, especially since there were no explicit directions on the OCAT assessment.

In fact, as students’ level of GO and TA increased, their level of engagement in free recall decreased. Students with a higher GO are known to engage in more SAL strategies (Tippin, Lafreniere, & Page, 2012). Since free recall relies more on DAL strategies, it makes sense that students with high GO are more inclined to skip this feature. Additionally, time constraints are a very significant contributor to students’ TA (Plass & Hill, 1986). The OCAT displays a 2-minute countdown clock with each new section, instead of the total 3-hour time limit. This time constraint may have made high-TA students more likely to skip sections in the OCAT assessment. While engagement in free recall decreased for those two constructs, it increased with greater levels of LO. Since high-LO students have more motivation and are more focused on mastering content through DAL, it makes sense that they are more engaged in free recall, compared to their lower LO counterparts. Students with higher LO also had higher levels of engagement in the immediate feedback feature. Therefore, these students actually took advantage of the explanations given to them and likely utilized the information to get a deeper understanding of the material. This engagement in DAL is typical of high-LO students (Race, 2005).

**Not always prompted: Second opportunity and retrieval cues.** Although free recall and immediate feedback are present for each multiple-choice question, the retrieval cue and second opportunity feature are not always prompted. They are prompted only when students get their first attempt at the question wrong. Thus, students may never actually engage in these two cognitive learning features. Having lower levels of self-reported engagement in these features would not be as problematic as the lack of engagement in free recall and immediate feedback. However, the second opportunity feature had the highest level of engagement, as students used it “almost every time.” Once students answered a question incorrectly the first time, they tried again for fewer marks. The student characteristics did not predict student engagement in the second opportunity feature, suggesting that they all engaged in it equally. It is understandable that students took advantage of this cognitive learning feature regardless of their orientation because they could still receive partial marks toward their final grade. Although this feature has grade-related benefits, it also has learning benefits that some students may have valued. For instance, the repetitiveness and ongoing retrieval needed in this feature help strengthen student knowledge and future retrieval processes. Therefore, the second opportunity feature can appeal to all students regardless of their academic orientation.

Since the second opportunity feature is paired with a retrieval cue, students also should have engaged in this feature “almost every time.” Instead, students indicated that they read or watched the hint that was prompted after the incorrect answer “only sometimes.” This suggests

that at times students are getting the answer wrong, skipping the retrieval cue, and answering the question again. In these cases, students may have been struggling between two multiple-choice options that had partial relevance. Once they realized that the first option was incorrect, they knew that their second option was the right answer, eliminating the need for a hint. Using a different assessment tool, Merrel, Cirillo, Schwartz, and Webb (2015) found similar results, with students continuously getting their second attempt correct. They concluded that students actually learned from their mistakes and were mastering the material. However, the retrieval cues provided in the OCAT are quite comprehensive and can help the students further understand the concept. Relearning the concept through the retrieval cues and immediate feedback can help strengthen students' future retrieval. Therefore, it was important for the students, regardless if they knew the right answer the second time, to actually take the effort to learn more about the concept.

Such avoidance of the retrieval cue was found in students with high GO. As students' level of GO increased, their level of engagement in the retrieval cue decreased. It was initially hypothesized that students high in GO would engage more in the retrieval cue because it acts as a hint, and therefore it can contribute to a higher grade. However, the retrieval cue requires students to take the time to read a portion of their textbook or watch a small video lecture. Such tasks are easy to avoid for GO students, who have more work avoidance and little motivation (Eison et al., 1986). In contrast, as students' level of LO increased, so did their engagement in the retrieval cues. Students with high LO are taking advantage of the retrieval cue and using it to get a deeper understanding of the material.

**Potential confounds to engagement.** Although student characteristics can act as an indicator for the limited engagement with the cognitive learning features, there are likely other factors. For instance, students may have already known the answer to the multiple-choice question and, in turn, found it counterproductive to engage in some of the cognitive learning strategies like free recall and immediate feedback. Additionally, students may be unaware of the educational and cognitive benefits of going through each of the features, which is a flaw for all assessment practices. Students perceive the grade as that final marker of success as opposed to the final marker being learning and the learning process. This again, demonstrates the disconnect the goal of the assessment (i.e., active engagement) and what is actually occurring (i.e., passive recognition). Finally, the students knew that they only received points for correct multiple-choice responses, regardless of engagement with the recall learning strategies. In other words, students did not receive any marks for answering the free recall portion of each question. Thus, students may have felt no incentive to engage with the other learning features because it had no immediate impact on their assessment grade. Students may have simply adapted their study strategies to meet the task demands. Thus, instead of reflecting a general lack of engagement, the findings could be the result of how students' study when they know they are only tested through multiple-choice instead of a mixture of recall and recognition questions. Nevertheless, bypassing the cognitive learning strategies in order to simply answer the question for a correct or incorrect response is not molding students into the active learners that they should be, which is the goal of this OCAT. In either situation, students did not appear to engage with the cognitive learning strategies implemented within the OCAT as originally intended by its developers.

## **Performance**

Students did extremely well on their OCAT exams, with an average of 90.6%, which differed significantly from the MCPP final exam average of 65.4%. Students' high marks on the OCAT assessment likely inhibited them from asking for help, providing help, or seeking advice

from their peers or professor. It may have created a false sense of security, but this was illusory, as no deep processing was occurring because they were skipping many of the cognitive learning features. This suggests that students need to engage with the cognitive learning features used in the OCAT to maximize their learning. Working through these problems in this way should be strengthening the learning process through DAL as opposed to simply memorizing and then recognizing, which is typical of MCPP tests. Additionally, if students were actually engaging in the cognitive learning strategies, that too should have contributed to mastery of the content. However, the problem is that if students were truly engaging in DAL, then that learning should have been translated into higher grades in the MCPP final exam, but it did not. For instance, Smith (2007) found that the students who frequently read their feedback on midterm assessments actually performed better on the summative assessment than their peers who never took the time to read it. Unfortunately, many of the students in the current study fell within that latter group, “only sometimes” reading the immediate feedback. As such, their final grades on the MCPP assessment suffered.

### **Limitations & Future Research**

There were several limitations to the current study, many of which were flaws in the methodological approach. First and foremost, attrition was extremely high from the presurvey to the postsurvey. This, in part, was a function of the distinct population of students who were at that time enrolled in the introduction to psychology course that was using the OCAT. Since only one instructor was implementing the assessment tool, the number of participants that could be recruited was limited. One reason why the presurvey–postsurvey design was initially chosen was that students needed to have enough exposure to the OCAT to answer the engagement questions. A larger sample size would have increased the statistical power in some of the regression analyses, ensuring greater confidence in the conclusions that were made and potentially uncover more statistically significant findings (Cohen, Cohen, West, & Aiken, 2003).

Additionally, engagement could have been measured more comprehensively, relying less on self-reports and more on the detailed diagnostic output from the OCAT. Not only does this detailed report indicate when the students start and finish each question, but it also provides reaction times for each cognitive learning feature. Also, all content that the students write in the free recall textbox can be viewed. In future research, free recall word counts and reaction times can be used as additional forms of engagement. Similarly, a program evaluation model, with researchers actually viewing students taking the assessment would also provide a more comprehensive look at this engagement process when using the OCAT. Having this comprehensive view would provide further and stronger evidence to the relationships that were found. Further, the regression models accounted for a limited amount of variance, suggesting that other factors are important in the predictive relationship for the outcomes of student engagement and performance. For instance, constructs such as student motivation, self-efficacy, and interest in course content should be included in future studies given their expected relationship.

Despite the aforementioned limitations, the current research still has very important implications, especially for the future implementation of the assessment tool. The most striking finding was the high OCAT grades, despite the limited engagement in the cognitive learning strategies. It calls into question whether students actually learned using the tool, especially since they performed significantly worse on their MCPP final exam. This may have resulted from their limited engagement with the OCAT. Students quickly become accustomed to the type of learning needed for success in classroom environments, recognizing that surface approach learning

strategies can be adequate for success in university. Implementing a new innovative assessment tool that integrates more deep approach learning strategies can be a difficult adjustment for students. This current study demonstrated that it is much easier for students to resist the change, resist the engagement, and resist the learning and instead find an easier, quicker way to receive a high grade. In the case of the OCAT being assessed, that meant skipping the built-in cognitive learning strategies, like free recall and immediate feedback, just to get to the multiple-choice options. Thus, because participants knew that they were only marked for correct multiple-choice options, we are unable to rule out that the students simply adapted their study strategies with this in mind.

### **Recommendation 1: Educate Students on the Cognitive Learning Strategies**

Cognitive learning strategies are there for a reason. Free recall, repetitive attempts, retrieval cues, and immediate feedback all have a strong theoretical background in enhancing long-term learning. If OCATs are being implemented in future academic classrooms, it is recommended that instructors or support staff alike educate the students on the built-in cognitive learning strategies for the tool. Communicating this educational piece about the OCAT and its corresponding cognitive learning features, would be an easy transition into the content that students already learn in their lectures on cognition. Additionally, the real-life application of this content can further motivate student learning (Gikandi, Morrow, & Davis, 2011; Kember, Ho, & Hong, 2008), and even enhance performance.

### **Recommendation 2: Use OCATs as a Studying Tool, Not Just Formal Assessment**

The high OCAT grades can, in part, be attributed to the second opportunity cognitive learning feature, which, when prompted, rewarded students with half a mark for a correct answer. These midterm averages increased the final course marks, despite the grades not being congruent with the final MCPP grades. This is a problematic indicator of student learning and can also pose problems for instructors meeting the institutional requirement of class averages for their course. As an alternative, instructors can provide OCAT as a studying tool for students, as opposed to a graded assessment. Despite the long-term learning benefits from practicing the cognitive learning strategies, students rarely have a formalized outlet for doing so. Instead, students are responsible for developing their own studying methods, which can be a grueling task for first year students who are not accustomed to assessments in higher education. Even further, it can help students with high GO and TA who are known to have greater difficulty in fostering strong and effective studying methods (Cassady & Johnson, 2002; Eison et al., 1986; Everson et al., 1995; Thomas, Cassady, & Heller, 2017).

### **Recommendation 3: Utilize Diagnostic Output to Mold Student Development**

Unlike other forms of assessment, the OCAT provides a very detailed diagnostic output for each individual student. Instructors have access to the students' grades, how many items they got correct on the first attempt, how many items they got correct on the second attempt, and how many items they got wrong. Even further, instructors can not only see the times when students started and finished the assessment, but also the times when they started and finished each individual question. Yet the most descriptive piece of information is students' reaction times to each stage of the assessment. These reaction times can provide insight to specific items that students had difficulty understanding, which can help mold student development in different content areas.

### **Conclusion**

Despite lack of student engagement with the OCAT, these tools should not be completely removed from an instructor's pedagogical toolbox. It was particularly beneficial for students with high LO and high TA, providing an outlet that harnessed their strengths and improved upon their weaknesses. However, the cognitive learning features are only meaningful when students actually take the time to use them to optimize their learning. As such, the OCAT assessed may have fallen victim to a cohort of students who are too focused on achieving that final grade and willing to bypass learning features that will help them get there. More research is needed to identify whether this type of behavior exists in other forms of online assessment. And if so, interventions on the importance of learning and cognition should be available to students.

### **Acknowledgements**

The authors would like to thank the OCAT developers Steve Joordens and Cho Kin Cheng for their brilliant work on their OCAT, their generosity in allowing us to use the online application in our classrooms, and their continued support on this research. We would also like to thank Dr. Kathryn Lafreniere, Dr. Geri Salinitri, and Elizabeth Ismaile for their many contributions to the project.

## References

- Baleni, Z. (2015). Online formative assessment in higher education: Its pros and cons. *The Electronic Journal of e-Learning*, 13(40), 228–236.
- Beck, H. P., Rorrer-Woody, S., & Pierce, L. (1991). The relations of learning and grade orientations to academic performance. *Teaching of Psychology*, 18(1), 35–37.
- Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research*, 40(1), 90–97.
- Cain, J., Romanelli, F., & Smith, K. M. (2012). Academic entitlement in pharmacy education. *American Journal of Pharmaceutical Education*, 76(10), 1–8.
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, 34, 268–276.
- Cassady, J. C. (2004). The impact of cognitive test anxiety on test comprehension and recall in the absence of external evaluative pressure. *Applied Cognitive Psychology*, 18(3), 311–325.
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27(2), 270–295.
- Chappuis, S., Stiggins, R., Arter, J., & Chappuis, J. (2004). *Assessment for learning: An action guide for school leaders*. Portland, OR: Assessment Training Institute.
- Coates, H. (2007). A model of online and general campus-based student engagement. *Assessment and Evaluation in Higher Education*, 32(2), 121–141.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- DeCaro, M. S., Thomas, R. D., Albert, N. B., & Beilock, S. L. (2011). Choking under pressure: Multiple routes to skill failure. *Journal of Experimental Psychology: General*, 140(3), 390.
- Dermo, J. (2009). E-assessment and the student learning experience. A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2), 203–214.
- Eison, J. A., Pollio, H. R., & Milton, O. (1983). Manual for use with LOGO-II. *Theory and Practice in Teacher Education and Other Works*. Retrieved from [http://trace.tennessee.edu/utk\\_theopubs/2](http://trace.tennessee.edu/utk_theopubs/2)
- Eison, J. A., Pollio, H. R., & Milton, O. (1986). Educational and personal characteristics of four different types of learning- and grade-oriented students. *Contemporary Educational Psychology*, 11, 54–67.
- Everson, H. T., Smodlake, I., & Tobias, S. (1995). Exploring the relationship of test anxiety and metacognition on reading test performance: A cognitive analysis. *Anxiety, Stress and Coping*, 7, 85–96.
- Frymier, A. B., & Weser, B. (2001). The role of student predispositions on student for instructor communication behaviour. *Communication Education*, 50, 314–326.

- Furlan, L. A., Cassady, J. C., & Pérez, E. R. (2009). Adapting the cognitive test anxiety scale for use with Argentinean university students. *International Journal of Testing, 9*(1), 3–19
- Furnham, A., Batey, M., & Martin, N. (2011). How would you like to be evaluated? The correlates of students' preferences for assessment methods. *Personality and Individual Differences, 50*, 259–263.
- Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*, 2333–2351.
- Greenberger, E., Lessard, J., Chen, C., & Farruggia, S. P. (2008). Self-entitled college students: Contributions of personality, parenting and motivational factors. *Journal of Youth and Adolescence, 37*, 1193–1204.
- Holbrook, J., Dupont, C., Power, M., & Joordens, S. (2015). *Enhancing learning through testing*. Presented at the OND Centre for Teaching Excellence Conference, Waterloo, Ontario. Retrieved from [https://uwaterloo.ca/centre-for-teaching-excellence/sites/ca.centre-for-teaching-excellence/files/uploads/files/mtuner\\_ond\\_2015\\_final.pdf](https://uwaterloo.ca/centre-for-teaching-excellence/sites/ca.centre-for-teaching-excellence/files/uploads/files/mtuner_ond_2015_final.pdf)
- Kalechstein, P., Hocevar, D., Zimmer, J. W., & Kalechstein, M. (1989). Procrastination over test preparation and test anxiety. In R. Schwarzer, H. M. van der Ploeg, & G. D. Spielberger (Eds.), *Advances in test anxiety research* (Vol. 6, pp. 63–76). Lisse, The Netherlands: Swets & Zeitlinger.
- Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science, 21*(3), 157–163.
- Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long term retention. *Journal of Memory and Language, 57*, 151–162.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968.
- Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language, 62*, 227–239.
- Kember, D., Ho, A., & Hong, C. (2008). The importance of establishing relevance in motivating student learning. *Active Learning in Higher Education, 9*, 249–263.
- Kopp, J. P., Zinn, T. E., Finney, S. J., & Jurich, D. P. (2011). The development and evaluation of the academic entitlement questionnaire. *Measuring and Evaluation in Counseling and Development, 44*(2), 105–129.
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989–998.
- Kuh, G. D., & Hu, S. (2001). The effects of student faculty interaction in the 1990s. *Review of Higher Education, 24*(30), 309–332.
- Lawrence, K. R., & Frymier, A. B. (2002). *The relationship of student-centered teaching with learning and grade orientation*. Paper presented at the annual meeting of the National Communication Association, New Orleans, LA.

- Levine, N. B. (2003). A multidimensional scaling analysis of perceptions of academic dishonesty among learning- and grade-oriented students. *Dissertation Abstracts International*, 64(4), 1175.
- Marsden, H., Carroll, M., & Neill, J. T. (2005). Who cheats at university? A self-report study of dishonest academic behaviours in a sample of Australian university students. *Australian Journal of Psychology*, 57(1), 1–10.
- Marsh, E. J., Roediger, H. L., Bjork, R. A., & Bjork, E. L. (2007). Memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, 14, 194–199.
- McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 11, 371–385.
- Merrel, J. D., Cirillo, P. F., Schwartz, P. M., Webb, J. A. (2015). Multiple-choice testing using immediate feedback-assessment technique forms: Second-chance guessing vs. second chance learning. *Higher Education Studies*, 5(5), 50–55.
- Miller, T. (2009). Formative computer-based assessment in higher education: The effectiveness of feedback in supporting student learning. *Assessment & Evaluation in Higher Education*, 34(2), 181–192.
- Noll, C. L., & Wilkins, M. (2002). Critical skills of IS professionals: A model for curriculum development. *Journal of Information Technology Education*, 1(3), 144–154.
- Oosterhof, A., Conrad, R. M., & Ely, D. P. (2008). *Assessing learners online*. New Jersey: Pearson.
- Özden, M. Y., Ertürk, I., & Sanli, R. (2004). Students' perceptions of online assessment: A case study. *Journal of Distance Education*, 19(2), 77–92.
- Pare, D. E., & Joordens, S. (2009). The invisible researcher: Using educational technologies as research tools for education. *Systemics, Cybernetics and Informatics*, 7(5), 37–42.
- Plass, J. A., & Hill, K. T. (1986). Children's achievement strategies and test performance: The role of time pressure, evaluation anxiety, and sex. *Developmental Psychology*, 22, 31–36.
- Purcell, J. M. (2010). Learning- and grade-orientations of community college students: Implications for instruction. *Community College Journal of Research and Practice*, 34, 497–511.
- Race, P. (2005). *Making learning happen*. London: Sage.
- Roedel, T. D., Schraw, G., & Plake, B. S. (1994). Validation of a measure of learning and performance goal orientations. *Educational and Psychological Measurement*, 54(4), 1013–1021.
- Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181–210.
- Saade, R. G., Morin, D., & Thomas, J. D. (2012). Critical thinking in e-learning environments. *Computers in Human Behaviour*, 28, 1608–1617.
- Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35, 453–472.

- Scouller, K., & Prosser, M. (1994). Student's experiences in studying for multiple choice question examinations. *Studies in Higher Education, 19*, 267–279.
- Smith, G. (2007). How does student performance on formative assessment relate to learning assessed by exams? *Journal of College Science Teaching, 36*(7), 28–34.
- Smith, S. D., & Caruso, J. B. (2010). *The ECAR study of undergraduate students and information technology* (ECAR Research Study 6). Boulder, CO: Center for Applied Research.
- Stöeber, J. (2004). Dimensions of test anxiety: Relations to ways of coping with pre-exam anxiety and uncertainty. *Anxiety, Stress & Coping, 17*(3), 213–226.
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education, 30*, 325–341.
- Tendhar, C., Culver, S. M., & Burge, P. L. (2013). Validating the national survey of student engagement (NSSE) at a research-intensive university. *Journal of Education and Training Studies, 1*(1), 182–193.
- Thomas, C. L., Cassady, J. C., & Heller, M. L. (2017). The influence of emotional intelligence, cognitive test anxiety, and coping strategies on undergraduate academic performance. *Learning and Individual Differences, 55*, 40–48.
- Tippin, G. K., Lafreniere, K., & Page, S. (2012). Student perception of academic grading: Personality, academic orientation and effort. *Active Learning in Higher Education, 13*(1), 51–61.
- Traub, R. E., & MacRury, K. (1990). Multiple-choice vs. free response in the testing of scholastic achievement. In K. Ingenkamp & R. S. Jager (Eds.) *Test und tens 8: jahrbuch der pädagogischen diagnostik* (Weinheim und Base, Beltz Verlag, pp. 128–159).
- Tulving, E. (1974). Cue-dependent forgetting. *American Scientist, 62*(1), 74–82.
- Walker, D. J., Topping, K., & Rodrigues, S. (2008). Student reflections on formative e-assessment: Expectations and perceptions. *Learning, Media and Technology, 33*, 221–234.
- Williams, K. D., & Frymier, A. B. (2007). The relationship between student educational orientation and motives for out-of-class communication. *Communication Research Reports, 24*, 249–256.
- Zeidner, M., & Matthews, G. (2005). Evaluation anxiety. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 141–163). London: Guildford Press.