



The importance of quality and quantity: An investigation into the influence of accuracy and frequency of retrieval practice on learning

Alice S. N. Kim^{1,3}, Linda Carozza² and Adam Sandford³

Abstract

There is a growing body of work on the application of cognitive learning principles to course design and instruction. In this study, we investigated the application of retrieval practice (retrieving information from memory) into the design of an online, asynchronous course. Specifically, we investigated whether: (1) the frequency; and (2) the accuracy of students' retrieval practice via practice quizzes was predictive of their final grade in the course in question. Both students' accuracy on the practice quizzes and the number of times they attempted these formative assessments were predictive of their final course grade, suggesting that both variables influence the benefit of retrieval practice on students' academic achievement in a course. These findings have important implications for course design and instruction, as they highlight the importance of encouraging students to engage in retrieval practice both regularly and with high accuracy.

Keywords

retrieval practice, testing effect, formative assessment, summative assessment, course design, cognitive learning principles

¹ Teaching and Learning Research In Action, Toronto, Canada

² York University, Toronto, Canada

³ University of Guelph-Humber, Toronto, Canada

Corresponding Author:

Alice S. N. Kim, Teaching and Learning Research In Action, Centre for Social Innovation,
720 Bathurst Street, Unit 501, Toronto, Ontario M5S 2R4

Email: alice.kim@tlraction.com

Introduction

A growing body of work on the application of cognitive learning principles has started to emerge in the literature (Dunlosky & Rawson, 2015; Higham et al., 2022; Rawson & Dunlosky, 2022). One principle that is particularly conducive to application in a classroom context is the testing effect, which in turn is linked to retrieval practice. The testing effect refers to the finding that retrieving information from memory (or retrieval practice) benefits knowledge retention (Agarwal et al., 2021; Roediger & Karpicke, 2006). Past research has also shown that the act of studying information repeatedly is unhelpful for long-term retention; instead, repeatedly retrieving information from memory has been shown to reinforce better what one has learned (Karpicke & Roediger, 2008). For the learner, the objective of studying information is to encode it so that it is accessible to them whenever they wish to retrieve it from memory, or in other words, so that they can retrieve the information from memory on demand.

More recent investigations have focused on implementing retrieval practice in the classroom to enhance students' learning and have shown that retrieval practice is beneficial for students at various stages of their studies (e.g., school age children to undergraduate and professional studies) and in a variety of disciplines (Dunlosky et al., 2013). Researchers are also advocating for the use of retrieval practice as a pedagogical strategy in course delivery and design (Agarwal & Bain, 2019; Dunlosky & Rawson, 2019; Weinstein et al., 2018), highlighting the importance of further investigation into factors that influence the successful implementation of retrieval practice into the design and instruction of a course. For course instructors who have limited time and resources, having a clear understanding of how to effectively implement retrieval practice in the classroom could help dramatically enhance students' learning.

Retrieval practice may be particularly useful for students' learning when it is completed through low-stakes, formative assessments that provide insight into students' current level of understanding of course material and gaps in their knowledge. These types of assessments are primarily intended to educate and improve students' academic performance (Wiggins, 1998), and to provide information about the extent to which students are achieving the intended course learning outcomes during the instruction of the course (Weston & McAlpine, 2004). In contrast, summative assessments are designed to communicate the extent to which students have grasped course learning objectives (Harlen, 2012) and they are intended to provide a record of students' academic achievement, reflecting the extent to which students have obtained knowledge associated with the course instruction (Gardner, 2010).

In this study, we investigated whether students': (1) performance on formative assessments that took the form of practice quizzes; and (2) the number of times they completed them are predictive of their final grade in the course. A better understanding of how these attributes of students' retrieval practice, namely quality and quantity, relates to their overall achievement in a course could help inform instructors on how to most effectively implement retrieval practice into the design and instruction of their course. Based on past studies, we hypothesized that the number of practice quiz attempts completed by students (quantity of retrieval practice) and their performance or accuracy on the practice quizzes (quality of retrieval practice) would be predictive of students' final grade in the course.

In the final week of the course, students were given the opportunity to earn bonus marks (2% on top of their final grade) by either: (1) agreeing to participate in this study, which effectively meant that they consented to having their coursework (including grades) used for research purposes; or (2) completing a reflective activity. All students who provided informed consent to participate in this study received the 2% bonus on top of their final grade irrespective of the number of times they completed each practice quiz (including whether they completed any at all) and how many questions they answered correctly on the practice quizzes. For the purposes of the analyses conducted for this study, the bonus marks were stripped from their final grades.

Method

Participants

The participants in this study consisted of students who were enrolled in an accelerated distance education course at a large Canadian university during the summer of 2019. The course was a general education course, and students were invited to participate in the present study during the last week of the course. Students were not told about the study beforehand. One hundred and seven (82%) of the students enrolled in the course provided informed consent to participate in the study. Of the 107 participating students, 46 were in their second year of a four-year degree, 46 were in their third year, 13 were in their fourth year, and two were in their first year (mean year of study = 2.65, S.D. = 0.72). Based on enrolment data, students who enrolled in this course were in various stages of their respective degrees and were majoring in a wide array of disciplines offered at the university. All students who participated provided informed consent to have their grades and coursework used for research purposes, following procedures approved by our institutional research ethics board.

Course

The 12-week course was instructed asynchronously online. Each week, students had a combination of reading material, lecture content, and a task-oriented activity to review and complete. Students also completed three major assignments outside of the weekly tasks. The weekly tasks alternated between participating in a team discussion forum and completing a quiz and were organized as a series of low stakes assignments that prepared students for the major assignments. Students were given an opportunity to think about and reflect on course ideas through the team discussion forum, and the quizzes provided students with an opportunity to practice applying knowledge learned as well as analyzing and evaluating argumentative discourse.

As described further below, students were evaluated in this course by way of participation, quizzes, a presentation, and two major writing assignments (an argumentative letter and a final critical essay). Participation was worth 10% of the final grade. Students had six opportunities to complete at least five participation posts that would be averaged to form their grade (five discussion posts at 2% each). Quizzes were worth 25% of the final grade. Students had six opportunities to complete at least five quizzes that would be averaged to form their grade (five quizzes at 5% each). Students had to present on a course concept. The visual presentation was asynchronous (e.g. lecture video, infographic, etc.) and worth 15% of the final grade. The remaining 50% of the grade was comprised of two writing assignments. An argumentative letter was worth 20% and a final critical essay was worth 30% of a student's final grade in the course. All

the grading for the course was completed by teaching assistants, none of whom are members of the research team. Since all quizzes were graded electronically via the course learning management system, the graders in this course did not have any knowledge of students' performance on the practice quizzes or the quizzes that contributed to students' final grades (only the instructor had access to this information).

Quizzes

Quizzes were embedded into the course learning management system, and they were completed by students in weeks 1, 3, 5, 7, 8, and 10. During these weeks, students learned the course material and had the opportunity to complete a practice quiz before completing the quiz that would be counted towards their final grade in the course. Since the course was asynchronous, students could decide when to take the quiz, provided it was completed during the relevant week. Once the week lapsed, then the relevant quizzes were closed. The questions asked on quizzes were predominantly in the form of multiple choice, with some true/false questions. Questions tested a combination of understanding, application of skills, and analysis (Bloom, 1956). Questions were predominantly derived from a repository of questions provided by the course's text publisher. Questions were also developed by the instructor. Questions and answer options on both the practice quizzes and actual quizzes were randomized by the learning management system. The quizzes that contributed to students' grades consisted of 10 to 12 questions, and these quizzes were timed. Students could only take these quizzes once. Quiz grades and feedback were provided only after the weekly window to complete the quiz had lapsed.

Practice quizzes

The practice quizzes consisted of approximately six questions (multiple choice and true/false), and one to two of these questions were included in the graded quiz. Students were not told that some of the questions on the practice quizzes would also be included on the graded quizzes, however, they may have become aware of this after they completed the first graded quiz. Practice quizzes were untimed and could be taken repeatedly until students were satisfied with their results, which were released as soon as the quiz was completed. The practice quizzes were implemented as optional formative assessments so that students had a means to assess their understanding of course material before taking a quiz (see Tessmer, 2013; Yorke, 2003). The feedback that students received for the practice quiz questions included a short explanation or text page that students could read to learn more about the content covered in the respective question.

Participation

Students were randomly divided into groups of approximately eight members. For bi-weekly participation tasks, each group in the course had a private discussion forum where they were given specific instructions for their post. Only the student group and the teaching team could view the written posts. The types of written posts students generated involved either application or analysis (Bloom, 1956). All assignments received a grade out of two marks. A typical high-level post was approximately 250 words. Students overall earned the full two marks. Exceptions to this were students who did not write enough or who repeated ideas of a team member who had already posted. Feedback was provided to each group. In most cases, one post was developed by the grader on the teaching team that related to the ideas of the team. In some instances, the grader commented on a particular student's post. This was always to the end of demonstrating a good idea or

making a connection with a student's post and the course material. Negative feedback was not provided on a team forum.

Argumentative letter

The argumentative letter was an independent written assignment between 750 and 900 words. This assignment was worth 20% of students' overall grade. It was completed in the first half of the course. Students were provided with the assignment instructions and given approximately a two-week deadline to upload their letters to the course learning management system. Students had three topics to choose from. The assignment required them to find a specific audience's stance on an issue (e.g. Elon Musk's stance on autonomous cars; Confino, 2024) and aim to convince this audience of an alternative perspective (e.g. autonomous cars would be detrimental for urban planning). Learning outcomes for this assignment included researching and summarizing arguments, providing counter-arguments, and addressing a resistant audience in an empathetic and collaborative manner as a technique of persuasion (also known as a Rogerian argument). In terms of Bloom's taxonomy, this involved the skills of understanding, evaluating, and creating (Bloom, 1956). Letters were graded by five different members of the teaching team.

Presentation

The presentation was worth 15% of the final grade. It was an independent assignment on the topic of a specific argument fallacy, and students presented their work to the same team members with whom they participated in a bi-weekly discussion forum. Students had to sign up for a specific argument fallacy in a wiki. No student worked on the same fallacy as another group member. Their task was to teach their group members about the fallacy by including a real-life example and providing an evaluation of it. The assignment was creative in nature, as students had discretion over how they presented their fallacy. Some students created animated videos, others made websites, blogs, etc. This presentation was evaluated based on demonstrating a combination of learning outcomes: understanding, application, evaluation, and creation (Bloom, 1956).

Final critical essay

The final critical essay was cumulative in nature. Students worked on it independently, and it was due after the term ended. It was between 1200 and 1400 words and worth 30% of students' final grade. Students were provided the assignment instructions and given approximately three weeks to upload their essays to the learning management system. Researching and writing this critical essay included: (1) choosing between two film options; (2) understanding the film's arguments; (3) analyzing all the arguments extensively by uncovering argument schemes and choosing which schemes to write about based on stronger analyses; (4) making connections with the schemes to other course concepts; (5) developing a critical thesis that tied together the analyses written about; and (6) writing the essay. The learning outcomes relevant to this writing assignment included diagramming, analysing, and evaluating (Bloom, 1956) arguments with methods and tools learned throughout the course. A student who earned a high grade (A/A+) neutrally provided a succinct analysis that was connected to a relevant critical thesis. The challenge of the essay was to avoid providing a list of analyses, but rather to make discretions and include strong analyses that were proven through the critical discussion of various arguments in the film.

Regression analysis

A multiple linear regression analysis was conducted to predict students' final grades based on: 1) their average number of practice quiz attempts; and 2) the average of their highest practice quiz grades.

Results

Table 1 present the means, standard deviations, maximum and minimum scores, and range for students': 1) average number of practice quiz attempts; 2) average of their highest practice quiz grades; and 3) final grade in the course for each student.

Table 1. Descriptive statistics for students': 1) average number of practice quiz attempts; 2) average of their highest practice quiz grades; and 3) final grade in the course for each student.

Variable	Mean	Standard deviation	Minimum score	Maximum score
No. of practice quiz attempts (average)	1.04	0.84	0	3.8
Highest practice quiz grade (average %)	66.58	13.97	28.51	95.9
Final grade in course (%)	71.02	10.21	16	87.01

Regression model

When the assumption of collinearity was tested, the results demonstrated that multicollinearity was not a concern (Tolerance = .993, VIF = 1.007). The data also met the assumption of independent errors (Durbin-Watson value = 1.904). A histogram of standardized residuals showed that the data contained approximately normally distributed errors, as did the normal P-P plot of standardized residuals, which showed points that were close to being on the line. The scatterplot of standardized predicted values indicated that the data met the assumptions of homogeneity of variance and linearity. The data also met the assumption of non-zero variances (average number of practice quiz attempts, Variance = .699; average highest practice quiz grade, Variance = 195.144; final grade in the course, Variance = 104.218).

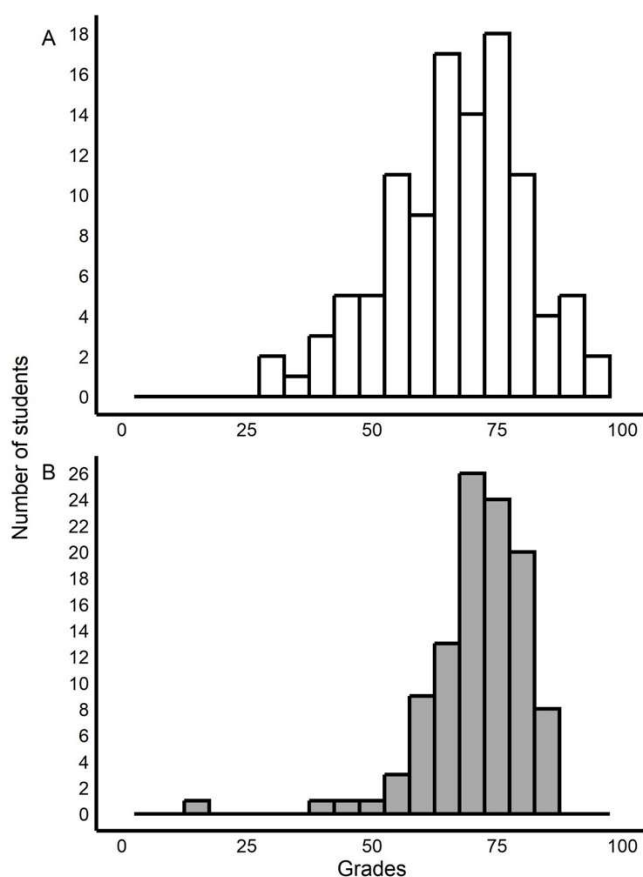


Figure 1. Histograms showing the distribution of average quiz grades (A) and final grades (B).

A significant regression equation was found ($F(2, 104) = 56.22, p < .001$), with an R^2 of .519. Participants' predicted final grade in the course is equal to $35.427 + .508$ (average of highest practice quiz grades) + 1.710 (average number of practice quiz attempts), where students' practice quiz grades and final grades in the course were measured in percentage and quiz attempts was measured in number. Students' predicted final grade in the course increased .508 percent for each percentage point of their average of highest practice quiz grades ($\beta = .695, t(104) = 10.188, p < .001$). Students' final grades increased by 1.71 percent for each increase in number of attempts ($\beta = .140, t(104) = 2.053, p = .043$).

Discussion

In this study, students were given the opportunity to complete practice quizzes as many times as they wanted within a predefined time frame. These practice quizzes were graded to provide students with corrective feedback, but they did not contribute to their final course grade. The practice quizzes provided a means of assessing the extent to which students engaged in retrieval practice opportunities provided by the instructor via formative assessments, while also keeping record of students' performance on these assessments. Based on past research, we hypothesized that both students' average number of practice quiz attempts (or quantity of quiz attempts) and the average of their highest practice quiz grades (or quality of quiz attempts) would be predictive of their final course grade, which was used to reflect their learning achievement in the course. Both variables were found to be predictive of students' final course grades. The regression model accounted for 52% of variance in final course grades, and quality of practice quiz

attempts accounted for greater variance in the model than quantity of attempts when units of measurement were standardized. In other words, engaging in retrieval practice is not the whole story; instead, performing well on these retrieval practice attempts also influences students' academic achievement in the course.

Our findings are in line with past research demonstrating the benefit of retrieval practice on knowledge retention in both lab-based and classroom settings (e.g., Agarwal et al., 2012; Agarwal et al., 2021; Carpenter et al., 2009; Carpenter et al. 2016; Sanchez et al., 2020). In this study, students' predicted final grades increased by approximately 2% for each additional increase in their averaged number of quiz attempts. However, our findings also suggest that the accuracy of one's retrieval attempts is important to their overall achievement in the course; students' predicted final grade in the course increased by about .5 percent for each additional percentage point added to the average of their highest practice quiz grades across all of the practice quizzes. These findings are consistent with past research demonstrating that retrieval practice increases the probability of being able to successfully retrieve information from memory but not the quality (or fidelity) of the information retrieved (e.g., Sutterer & Awh, 2016). This may explain why both the quality of students' retrieval practice (as reflected by their grades on the practice quizzes) and the quantity of retrieval practice they engaged in (reflected by their average number of attempts across each of the practice quizzes) were significant predictors of students' final course grades.

Our findings are also consistent with past research demonstrating that students' performance on formative assessments are predictive of their final course grades (Connor et al., 2006; Jensen & Barron, 2014), and that completing formative assessments influences students' achievement in the course (Carrillo-de-la-Peña et al., 2009). In the present study, practice quizzes were both completely optional for students to complete and were not counted towards their grade in the course, though students may have come to realize that up to two questions on the practice quizzes were also included on the quizzes that counted towards their final grade. Students' final grades were calculated based on various assignments, as described above, that took on various forms. Although the format of the quizzes was similar to that of the practice quizzes, this was not the case for any of the other assignments, including the presentation, argumentative letter, and final critical essay; this suggests that formative assessments that are intended to engage students in retrieval practice to prepare students for the summative assessments do not necessarily need to take on the same format as the summative assessments. As mentioned above, in the present study we found that the accuracy of students' retrieval practice was predictive of their grades on all of the individual course assessments, and we interpret these results to suggest that students were indeed learning and not just memorizing the relevant course material. Future research should investigate whether the format of formative and summative assessments influence the benefit of retrieval practice on students' learning.

The findings of our study have implications for course design and instruction. More specifically, our findings suggest that students would benefit from engaging in frequent and high-fidelity retrieval practice that is integrated into the course design. Instructors can encourage their students to engage in this practice by incorporating practice quizzes and other formative assessments into the course and encouraging their students to perform well on them. This may include assigning a small portion of students' final grades to these retrieval practice opportunities based on their performance and emphasizing the

importance of performing well on these low stakes assignments as an effective means of preparing for summative assessments that account for more of their final course grade. Additionally, incentivizing students' participation in class activities through gamification has become a growing trend (Göksün & Gürsoy, 2019). For example, digital badges have been distributed to students based on participation (participatory badges) or achievement of course-related competencies (performance badges), and it has been suggested that using these digital badges engages students' motivation to learn, recognition of skill development, and/or interaction with others (e.g., Chou & He, 2017; Shields & Chugh, 2017). However, past research demonstrates mixed findings regarding the influence of digital badging on students' learning (for a systematic review, see Roy & Clark, 2019).

With respect to the outcome of using traditional quizzes (i.e., multiple choice questions) vs. gamified online quizzes (i.e., an option to place a wager) as a means for students to prepare for tests, Sanchez et al. (2020) demonstrated that students' performance on tests improved as they completed more quizzes irrespective of quiz format (traditional vs. gamified), which aligns with the findings of the present study. In addition to the number of quiz attempts, however, our results also show that students' performance on the practice quizzes was predictive of their final course grades, which is novel relative to Sanchez et al.'s (2020) findings. Regarding using gamification as a means of encouraging students to engage in frequent and high-fidelity retrieval practice, Sanchez et al. found that students in the gamified quiz group did not complete more quizzes than students in the traditional quiz group. Additionally, although students in the gamified quiz group were found to score higher on the first test, this did not continue for the subsequent tests, leading the researchers to suggest that the benefit of gamification may occur through a novelty effect and for this reason may not be sustainable. However, the influence of gamification on student learning may depend on the details of how gamification is used. For example, an online social network with gamification elements was shown to successfully encourage and motivate students to complete optional online quizzes (Landers & Callan, 2011). Thus, although the extant literature on the influence of gamification on student learning presents mixed findings, future work should tease apart the different ways that gamification elements are used in education and how they influence students' learning.

The current study focused on whether the quantity and quality of retrieval practice attempts via practice quizzes could predict final course grades across a sample of undergraduate students. Past studies have found that individual differences (in variables such as intelligence, motivation, conscientiousness, and self-efficacy) predict academic success (e.g., Farsides & Woodfield, 2003; Feldman & Kubota, 2015; O'Connor & Paunonen, 2007; Steinmayer & Spinath, 2009; Vedel, 2014; Zuffianò et al., 2013). For example, in one meta-analysis including 17,717 participants and 21 correlations between Big Five personality dimensions and grade point average found that conscientiousness was the most predictive of academic success (Vedel, 2014). These, or other, individual differences could interact with the number of attempts and/or performance on practice quizzes, thereby moderating their predictive influence on academic success. For example, more conscientious students could either attempt more practice quizzes or perform better on practice quizzes by preparing for them than less conscientious students. Future research should explore whether individual differences influence engagement with practice quizzes.

The findings of the present study are specific to a particular learning context, that is, a distance education course in Philosophy (general education) that was offered in Ontario, Canada before the COVID-19 pandemic when remote, asynchronous instruction became much more common. Consequently, the results of the present study should be generalized to other learning contexts with caution. Additionally, further work should be done to differentiate any influence of practice effects from the number of times students completed a practice quiz (or engaged in retrieval practice).

The findings of the present study shed light on how instructors can encourage their students to engage in effective learning strategies, specifically retrieval practice, by providing their students with structured opportunities to do so as part of the design of the course. Additionally, our findings suggest that students may benefit from: (1) engaging in retrieval practice frequently; and (2) trying their best to retrieve the sought-after information accurately from memory. Future research should investigate the following topics related to retrieval practice: (1) the efficacy of using various gamification components to help encourage students to engage in high-fidelity retrieval practice; (2) the influence of individual factors that may moderate the effects of retrieval practice; as well as (3) whether the alignment between the format of formative assessments used to elicit retrieval practice and the summative assessments that students are preparing for influence students' performance on the summative assessment.

References

- Agarwal, P. K., & Bain, P.M. (2019). *Powerful teaching: Unleash the science of learning*. Jossey-Bass. <https://doi.org/10.1002/9781119549031>
- Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist. *Educational Psychology Review*, 24(3), 437–448. <https://doi.org/10.1007/s10648-012-9210-2>
- Agarwal, P.K., Nunes, L.D. & Blunt, J.R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review*, 33, 1409–1453. <https://doi.org/10.1007/s10648-021-09595-9>
- Bloom, B. S. (1956). *Taxonomy of educational objectives* (Vol. 1: Cognitive domain, pp. 20–24). McKay.
- Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, 23(6), 760–771. <https://doi.org/10.1002/acp.1507>
- Carpenter, S. K., Lund, T. J. S., Coffman, C. R., Armstrong, P. I., Lamm, M. H., & Reason, R. D. (2016). A classroom study on the relationship between student achievement and retrieval-enhanced learning. *Educational Psychology Review*, 28(2), 353–375.
- Carrillo-de-la-Pena, M. T., Bailles, E., Caseras, X., Martínez, À., Ortet, G., & Pérez, J. (2009). Formative assessment and academic achievement in pre-graduate students of health sciences. *Advances in Health Sciences Education*, 14(1), 61–67. <https://doi.org/10.1007/s10459-007-9086-y>
- Chou, C. C., & He, S.-J. (2017). The effectiveness of digital badges on student online contributions. *Journal of Educational Computing Research*, 54(8), 1092–1116. <https://doi.org/10.1177/0735633116649374>
- Confino, P. (2024, April 23). On a crucial earnings call, Musk reminds the world Tesla is a tech company. 'Even if I'm kidnapped by aliens tomorrow, Tesla will solve autonomy'. *Fortune*. <https://fortune.com/2024/04/23/tesla-elon-musk-electric-vehicles-waymo-uber/>

- Connor, J., Franko, J., & Wambach, C. (2006). *A brief report: The relationship between mid-semester grades and final grades*. University of Minnesota, USA.
- Dunlosky, J., & Rawson, K. A. (2015). Practice tests, spaced practice, and successive relearning: Tips for classroom use and for guiding students' learning. *Scholarship of Teaching and Learning in Psychology*, 1(1), 72–78. <https://doi.org/10.1037/stl0000024>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14, 4–58. <https://doi.org/10.1177/1529100612453266>
- Dunlosky, J., & Rawson, K. A. (Eds.) (2019). *The Cambridge handbook of cognition and education*. Cambridge University Press. <https://doi.org/10.1017/9781108235631>
- Dunlosky, J., & Rawson, K. A. (2015). Practice tests, spaced practice, and successive relearning: Tips for classroom use and for guiding students' learning. *Scholarship of Teaching and Learning in Psychology*, 1(1), 72–78. <https://doi.org/10.1037/stl0000024>
- Gardner, J. (2010). Developing Teacher Assessments: An Introduction. In J. Gardner, W. Harlen, L. Hayward, G. Stobart, & M. Montgomery (Eds.), *Developing Teacher Assessment* (pp.1-11). McGraw-Hill/Open University Press.
- Farsides, T., & Woodfield, R. (2003). Individual differences and undergraduate academic success: The roles of personality, intelligence, and application. *Personality and Individual Differences*, 34(7), 1225-1243. [https://doi.org/10.1016/S0191-8869\(02\)00111-3](https://doi.org/10.1016/S0191-8869(02)00111-3)
- Feldman, D. B., & Kubota, M. (2015). Hope, self-efficacy, optimism, and academic achievement: Distinguishing constructs and levels of specificity in predicting college grade-point average. *Learning and Individual Differences*, 37, 210-216. <https://doi.org/10.1016/j.lindif.2014.11.022>
- Göksün, D. O., & Gürsoy, G. (2019). Comparing success and engagement in gamified learning experiences via Kahoot and Quizizz. *Computers & Education*, 135, 15–29. <https://doi.org/10.1016/j.compedu.2019.02.015>
- Harlen, W. (2012). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (pp. 87-102). SAGE Publications Ltd. <https://www.doi.org/10.4135/9781446250808.n6>
- Higham, P. A., Zengel, B., Bartlett, L. K., & Hadwin, J. A. (2022). The benefits of successive relearning on multiple learning outcomes. *Journal of Educational Psychology*, 114(5), 928–944. <https://doi.org/10.1037/edu0000693>
- Jensen, P. A., & Barron, J. N. (2014). Midterm and first-exam grades predict final grades in biology courses. *Journal of College Science Teaching*, 44(2), 82-89. https://doi.org/10.2505/4/jcst14_044_02_82
- Karpicke, J. D., & Roediger, H. L., 3rd. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <https://doi.org/10.1126/science.1152408>
- Landers, R.N., & Callan, R.C. (2011). Casual social games as serious games: The psychology of gamification in undergraduate education and employee training. In: Ma, M., Oikonomou, A., Jain, L. (eds) *Serious games and edutainment applications*. Springer. https://doi.org/10.1007/978-1-4471-2161-9_20
- Liyanagunawardena, T. R., Scalzavara, S., & Williams, S. A. (2017). Open badges: A systematic review of peer-reviewed published literature (2011-2015). *European Journal of Open, Distance and E-Learning*, 20(2), 1-16. <https://doi.org/10.1515/eurodl-2017-0013>

- O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, 43(5), 971-990. <https://doi.org/10.1016/j.paid.2007.03.017>
- Rawson, K. A., & Dunlosky, J. (2022). Successive relearning: An underexplored but potent technique for obtaining and maintaining knowledge. *Current Directions in Psychological Science*, 31(4), 362-368. <https://doi.org/10.1177/09637214221100484>
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249-255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Roy, S., & Clark, D. (2019). Digital badges, do they live up to the hype? *British Journal of Educational Technology*, 50, 2619-2636. <https://doi.org/10.1111/bjet.12709>
- Rutter, A., (2020). Fully asynchronous online courses: What works for student success. *Teaching, Learning & Assessment*, 41. <https://digscholarship.unco.edu/tla/41>
- Sanchez, D.R., Langer, M., & Kaur, R.D. (2020). Gamification in the classroom: Examining the impact of gamified quizzes on student learning. *Computers & Education*, 144. <https://doi.org/10.1016/j.compedu.2019.103666>
- Shaw, V. N. (2001). Training in presentation skills: An innovative method for college instruction. *Education*, 122(1), 140-144.
- Shields, R., & Chugh, R. (2017). Digital badges – rewards for learning? *Education and Information Technologies*, 22, 1817-1824. <https://doi.org/10.1007/s10639-016-9521-x>
- Steinmayer, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, 19(1), 80-90. <https://doi.org/10.1016/j.lindif.2008.05.004>
- Sutterer, D.W., & Awh, E. (2016). Retrieval practice enhances the accessibility but not the quality of memory. *Psychonomic Bulletin & Review*, 23, 831-841 (2016). <https://doi.org/10.3758/s13423-015-0937-x>
- Tessmer, M. (2013). *Planning and conducting formative evaluations*. Routledge. <https://doi.org/10.4324/9780203061978>
- Vedel, A. (2014). The Big Five and tertiary academic performance: A systematic review and meta-analysis. *Personality and Individual Differences*, 71, 66-76. <https://doi.org/10.1016/j.paid.2014.07.011>
- Weinstein, Y., Madan, C. R., & Sumeracki, M. A. (2018). Teaching the science of learning. *Cognitive Research: Principles and Implications*, 3, 2. <https://doi.org/10.1186/s41235-017-0087-y>
- Weston, C., & McAlpine, L. (2004). Evaluation of student learning. In A. Saroyan & C. Amundsen (Eds.) *Rethinking teaching in higher education: From a course design workshop to a faculty development framework*. Stylus publishing. <https://doi.org/10.1111/j.1467-9647.2006.00291.x>
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(4), 477-501. <https://doi.org/10.1023/A:1023967026413>
- Zuffianò, A., Alessandri, G., Gerbino, M., Kanacri, B. P. L., Guinta, L. D., Milioni, M., & Caprara, G. V. (2013). Academic achievement: The unique contribution of self-efficacy beliefs in self-regulated learning beyond intelligence, personality traits, and self-esteem. *Learning and Individual Differences*, 23, 158-162. <https://doi.org/10.1016/j.lindif.2012.07.010>