Journal of Scientific Psychology
May 2025
Volume 20
Pages 33 - 41

Using the Immediate Feedback Assessment Technique (IFAT) for Test Reviews: Effects on Academic Performance

Lynne N. Kennette,[1] Phoebe S. Lin, [2] & Lisa R. Van Havermaet[3]

[1] *Durham College, Ontario, Canada*
[2] *Framingham State University, Massachusetts, United States*
[3] *University of Michigan, Michigan, United States*

Abstract

Utilizing a variation of the Immediate Feedback Assessment Technique (IFAT) activity during weekly reviews, the present study examined combined effects of collaboration, retrieval, and immediate feedback to determine whether it affected student performance in the class. Participants in the control group did not use the IFAT while those in the two experimental conditions used it either for the full semester or began using it only partway through the semester. Hypotheses were partially supported. The benefit of using the IFAT was demonstrated as both IFAT conditions (part-semester and full semester) obtained higher grades than the control group. However, we did not find an incremental benefit as we had hypothesized where using the IFAT more would lead to more gains than using it for part of the semester. Now that most students are back to in-person learning, educators are encouraged to consider using the IFAT to optimize student success.

Keywords: performance, immediate feedback, collaboration, retrieval, immediate feedback assessment technique

Using the Immediate Feedback Assessment Technique (IFAT) for Test Reviews: Effects on Academic Performance

Many studies have shown us that processes such as rehearsal are able to move information from short-term memory into a more permanent long-term memory storage (e.g., Bird et al., 2015; Himmer et al., 2019; Squire et al., 2015). Current evidence supports the conclusion that being exposed to information multiple times (especially in different contexts) will facilitate its movement into long-term memory storage (Baddeley, 1997). Others have identified that deeper processing of information (e.g., by elaborating, engaging with it in a meaningful way, and/or connecting it to information which is already known) also helps information get stored in a way that facilitates later retrieval (Craik & Lockhart, 21972; Baddeley & Hitch, 2017; Bjork, 2011; Weinstein et al., 2018). In addition, we have also learned a great deal about the effects of retrieval, collaboration (including team-based learning), and immediate feedback, which all appear to offer a significant contribution to improving memory. The present study seeks to explore the effects of a modified IFAT (Immediate Feedback Assessment Technique) activity which combines all of these processes: retrieval, collaboration, and immediate feedback.

Retrieval

In a classic study, Roediger and Karpicke (2006) demonstrated that retrieval practice is far better for future memory performance than rehearsal. They asked participants to read about a topic such as the sun and then answer some questions about the information they read to assess retention. They then asked some participants to study the material by re-reading the passage (called a study block) and others to write down everything they could recall reading about the topic (called a test block). In both cases (test and study blocks), the participants engaged in the

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Lynne N. Kennette, Durham College, 2000 Simcoe St N. Oshawa, Ontario, Canada L1G 0C5 (905)721-2000 ext. 3704 lynne.kennette@durhamcollege.ca

activity for the same amount of time, and retention was assessed 5 minutes later, 2 days later, and 1 week later. Although studying demonstrated better retention after the 5-minute delay, those who completed the retrieval activity (test block) significantly out-performed the students who re-read the passage (study block) after a longer delay (2 days and 1 week). In a second experiment, they replicated these findings demonstrating the incremental advantage of testing over studying where increasingly better retention was shown as the number of study blocks increased. This confirmed that testing is a superior strategy for long term memory retention. Retrieving information (i.e., retrieval) from memory produces superior long-term retention compared to re-reading information (i.e., rehearsal). This conclusion was supported more recently in similar studies (e.g., Arial & Karpicke, 2018; Butler, 2010) and in a meta-analysis by Adesope et al. (2017).

Collaboration

The literature examining team-based learning generally finds that this type of collaborative learning approach is beneficial to student learning (Davidson et al., 2014; Fink, 2004). For example, in a two-stage collaborative testing paradigm, Cortright et al. (2003) demonstrated that students who were allowed to complete the test a second time in small groups out-performed their peers on that content when re-tested on a final exam four weeks later. This demonstrates that collaboration can improve long-term retention.

In another example, Rajaram and Pereira-Pasarin (2007) showed that those permitted to engage in a group discussion prior to testing out-performed those who were not given this opportunity. In their study, after learning some words and pictures, some participants were able to discuss their answers in a group of three before completing a recognition test. The performance benefit, experienced by the group that was able to discuss in a group, appeared to be long-lasting, as it was shown after 1 hour, 48 hours, and 1 week.

Another benefit of a group discussion is that stronger students may be able to provide better explanations of concepts to weaker students on account of their more similar cognitive networks compared to an expert instructor (see Bowman et al., 2013), and teaching someone else about a topic (e.g., in the case of a tutor) has also shown retention advantages for the "teacher" (Fiorella & Mayer, 2013, 2014; Hoogerheide et al, 2016). The benefits of collaborative learning have been well-documented (see Laal & Ghodsi, 2011 and Nokes-Malach, Richey, & Gadgil, 2015 for a thorough review). A number of underlying mechanisms (both cognitive and social) have been proposed to be involved in the benefit seen for collaboration including increased working memory resources, error-correction, reexposure, relearning through retrieval, observational learning, increased motivation, and cross-cueing (Nokes-Malach et al., 2015).

Immediate Feedback

Long-term learning is also improved when we are given immediate feedback about our performance. For example, Dihoff et al. (2004) demonstrated that feedback improved retention (24 hours after the test, immediately after the entire test was completed, or immediately after each question) compared to receiving no feedback. Similarly, Pashler et al. (2005) varied the feedback received by participants (no feedback, "correct"/ "incorrect" only, or shown the correct answer) and found that immediate feedback indicating the correct response greatly improved retention when tested one week later. One mechanism that could explain why immediate feedback is beneficial is that it could allow for immediate verification and elaboration, or prevent the storage of incorrect information (Shute, 2008). Therefore, educators can incorporate feedback into the teaching and learning process to improve student performance.

Recently, Kennette and McGuckin (2018) have shown that using the IFAT to review course content throughout the semester resulted in higher final grades compared to a control group, but that this effect was driven primarily by assignment grades. Here, we look to replicate and extend their finding and examine the effect of partial IFAT use (for only part of the semester) on overall student performance examining both assignment and test grades.

Hypotheses

We expected to find an incremental benefit to using the IFAT to review course content due to the benefits of collaboration (team-based learning), spaced retrieval, and immediate feedback, which will result in these performance differences. Specifically, in the group with no IFAT use, we expected the lowest performance (final grades, test scores, and assignment scores) and we expect to find the highest scores in the group which will use the

IFAT through the entire semester. The performance is expected to fall somewhere in between for the group which used the IFAT during only half of the semester.

Method

Participants

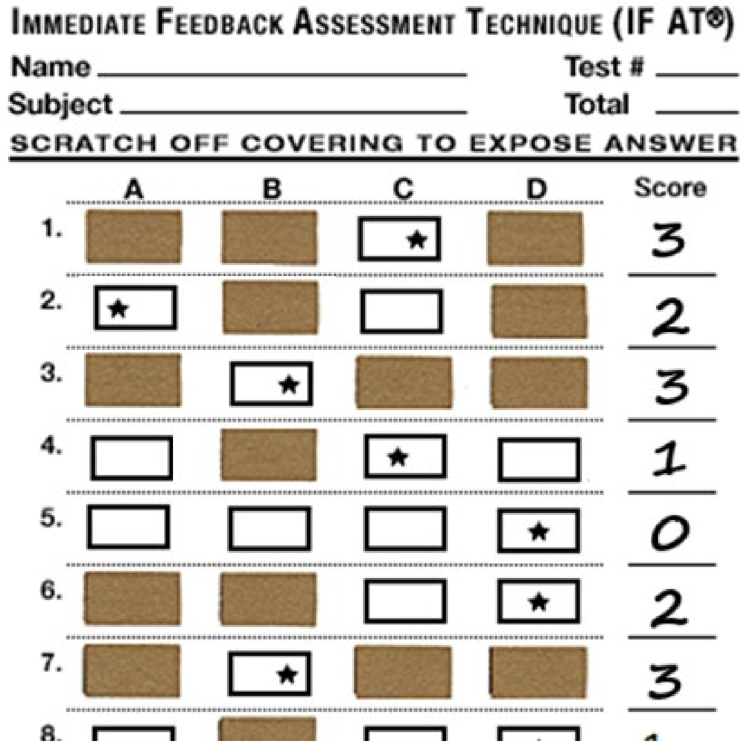
Data were collected from three groups of Canadian community college students (across 3 semesters) enrolled in a general education psychology course (*n* = 36, *n* = 42, *n* = 55). Since demographic variables were not hypothesized to impact differences in performance, this information was not collected in order to protect student privacy and ensure anonymity. Generally speaking, however, students were from a variety of programs and enrolled in different sections of the general education course Introduction to Psychology, which was taught by the same instructor.

Materials and Procedure

The Immediate Feedback Assessment Technique (IFAT) is a tool developed by Epstein Education (which can be purchased through CognaLearn at https://www.cognalearn.com/IFAT). It is a small paper card which resembles a scantron form in that there are question numbers listed at the left and 4 boxes (labelled A, B, C, and D) to the right of each question where students indicate their response to each question by filling in the corresponding bubble. It is different from a scantron form, however, in that instead of filling in the boxes to indicate their answer choice, students scratch off the box of their answer to reveal a star (if correct) or nothing (if incorrect). Students therefore receive immediate feedback on their performance and can continue to narrow down the correct answer (see Figure 1). Interested readers can read all about the ways that cheating is deterred in the design of this card (e.g., the star could be in any location within the box, there are black bands printed on the back of the card so that if students hold it up to the light, the star cannot be seen, etc.). The instructor, assignments (and rubrics), and test questions were identical across the three groups and the instructor used the same slides, in-class activities, and examples as well for standardization purposes.

In one section of the course, students did not use the IFAT as part of a weekly review of the course material. In a second section of the course, the IFAT was introduced part-way through the semester as a way to review course content weekly. And in a third section, the IFAT was used weekly from the beginning of the semester. When the IFAT was used (i.e., in both the second and third groups), students formed small groups of 3-5 students in class and collaborated to complete the 10 assigned multiple choice questions using a single IFAT scoring card. In this way, groups needed to discuss and collaborate in order to select the correct response and answer using the IFAT card (by scratching the correct square). Students were discouraged from using their notes or other sources of information to answer these questions. Instead, they were instructed to rely only on their group members and those group discussions to identify the correct answer; if their first selection was incorrect, they proceeded to identify the next most likely correct answer. The IFAT review activity was the first activity presented during the weekly 3-hour class and reviewed the previous week's material. As an incentive for this weekly activity (since it was not tied directly to marks), the instructor provided a sweet treat such as a piece of candy or small chocolate bar to the group with the highest score every week. The three groups were compared on their performance during the semester: average assignment grades (2 assignments total), average test scores (3 tests total), and final grades at the end of the semester. This project was approved by the institution's Research Ethics Board prior to commencement. Because this project analyzed data collected in prior semesters for another purpose, the Research Ethics Board confirmed that the data should be treated as secondary use research and therefore informed consent was not explicitly obtained for these data, conforming to the ethical standards of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans- TCPS2 (Canadian Institutes of Health Research, 2018) which governs research ethics in Canada.

Figure 1
Image of the Immediate Feedback Assessment Technique card (adapted from http://testmaker.if-at.com/home/demo/demo1.htm)



Results

Analyses utilized one-way analysis of variance (ANOVA) with Tukey's post hoc tests to compare no IFAT use, partial IFAT use, and full IFAT use to determine if this review strategy affected academic performance. Initial analyses compared the three conditions of IFAT use to determine the effects on average assignment grades, average test scores, and final grades at the end of each semester. Descriptive statistics are in Table 1.

Table 1
Descriptive Statistics Indicating Significant Differences Across Conditions

	Mean	SD
Average Assessment Grade		
No IFAT use	80.82	10.69
Partial IFAT use	90.35	9.27
Full IFAT use	82.95	10.90
Average Test Grade		
No IFAT use	71.82	13.69

Partial IFAT use	82.57	11.84
Full IFAT use	76.96	12.48
First Assignment		
No IFAT use	75.95	15.94
Partial IFAT use	90.29	11.69
Full IFAT use	80.92	12.56
Second Assignment		
No IFAT use	85.68	14.03
Partial IFAT use	90.41	12.07
Full IFAT use	84.98	13.85
First Test		
No IFAT use	74.15	16.87
Partial IFAT use	82.36	14.79
Full IFAT use	79.48	11.96
Second Test		
No IFAT use	71.31	14.41
Partial IFAT use	81.09	12.71
Full IFAT use	76.26	14.21
Third Test		
No IFAT use	70.00	15.56
Partial IFAT use	84.24	14.34
Full IFAT use	75.15	16.35

Average Assignment Grades

To analyze whether students differed in assignment scores across conditions, any students who had not submitted one or more assignment (evidenced by an assignment grade of zero) were omitted from the analyses with the final number of participants in the secondary set of analyses being 108 participants (22 in the no IFAT condition, 34 in the partial IFAT condition, and 52 in the full IFAT condition). In terms of average assignment grades, students in the partial IFAT condition ($M = 90.35$, $SD = 9.27$) significantly outperformed both the no IFAT condition ($M = 80.82$, $SD = 10.69$) and the full IFAT condition ($M = 82.95$, $SD = 10.90$), $F(2, 105) = 7.32$, $p < 0.01$.

Average Test Grades

Similarly, students who earned a zero on any of the three tests in the course were excluded from the analysis comparing average test scores, with the number of participants (*N* = 113) in the sample: 26 in the no IFAT condition, 33 in the partial IFAT condition, and 54 in the full IFAT condition. Although the sample sizes were unequal, this small variation should not pose a problem for these analyses, as they are based on group means (Wickens & Keppel, 2004). In terms of the mean score across the three tests, those in the partial IFAT condition (*M* = 82.57, *SD* = 11.84) significantly outperformed those in the no IFAT condition (*M* = 71.82, *SD* = 13.69), $F(2, 110) = 5.36$, $p < 0.01$.

Final Grades

Results did not indicate significant differences across groups in terms of final grades $F(2, 99) = 2.11$, $p = 0.13$, however, the means were directionally consistent with aforementioned findings such that those in the partial IFAT condition (*M* = 84.37, *SD* = 10.76) obtained higher scores than those in the no IFAT (*M* = 78.40, *SD* = 9.70) and full IFAT (*M* = 81.62, *SD* = 9.84) conditions.

Individual Components of Assessments

To further probe the above effect of the IFAT on performance, additional analyses were conducted using individual scores on each of the assignments separately and the three tests individually as dependent variables. Analyses indicated significant differences on grades for the first assignment, second test, and third test.

Assignments

Results indicated that on the first assignment, students in the partial IFAT condition (*M* = 90.29, *SD* = 11.69) earned significantly higher scores than those in the no IFAT condition (*M* = 75.95, *SD* = 15.94) and the full IFAT condition (*M* = 80.92, *SD* = 12.56), $F(2, 105) = 9.17$, $p < 0.001$. No significant differences across groups were found in terms of performance on the second assignment.

Tests

No significant differences across groups were found for performance on the first test. Significant differences, however, were found for test performance on the second and third tests. Results indicated that the students in the partial IFAT condition (*M* = 81.09, *SD* = 12.71) significantly outperformed those in the no IFAT condition (*M* = 71.31, *SD* = 14.41) on the second test, $F(2, 110) = 3.66$, $p < 0.05$. Similarly, students in the partial IFAT condition (*M* = 84.24, *SD* = 14.34) earned significantly higher scores on the third test compared to those in the no IFAT condition (*M* = 70.00, *SD* = 15.56) and those in the full IFAT condition (*M* = 75.15, *SD* = 16.35), $F(2, 110) = 6.51$, $p < 0.01$.

Discussion

Our results indicate that the partial use of the IFAT to review content resulted in better performance than not using it, but this partial use also sometimes outperformed the full IFAT condition (on the third test, on their average score across all 3 tests, on assignment 1, and on their average assignment score across both assignments).

Although designed by Epstein Education to replace scantron forms during testing, the present study employed the IFAT cards here as a tool to use during weekly review of the course content. The "answer until correct" procedure has been shown to benefit students' learning and provide partial credit for partial knowledge during testing (in fact the IFAT has also been named to reflect the first part of the saying: *"if at" first you don't succeed, try, try again*), among other benefits (Chang & Li, 2019; Slepkov & Godfrey, 2019).

It is quite intuitive to explain the difference found in partial support of our hypotheses. Notably, that using the IFAT would lead to improved performance over not using it at all. In all cases, using the IFAT to review material at least for part of the semester resulted in better performance on a number of outcome measures. However, the partial-use group consistently outperformed the group that used the IFAT for the entire semester. We did not expect this flipped direction, as we had predicted the best outcome for full-semester use. In almost every case, however, the no IFAT condition yielded the lowest scores (the exception being Assignment 2). One possible explanation is that there could have been a priori differences between the three groups of students. Since no demographic data were collected and no baseline measure exists in our dataset (e.g., overall GPA), we cannot know for certain whether the three classes used were the same at the start of the semester, or whether they varied in important ways that could explain these differences. We have no reason to expect them to be different, however, as students who register for the course are more or less randomly selecting a section of the course.

It is interesting to note that in every case, the group in the partial IFAT condition had less variation in their scores (having the smallest SD), so it could be that there was indeed something different about this group compared to the other two. There could also be additional variables which moderated this effect, but which were not measured here. One such example is working memory (WM) capacity. Agarwal et al. (2017) showed that students with lower working memory capacity benefitted more from a retrieval practice intervention. So, it could be the case that the students in the partial IFAT condition had lower working memory capacity, and so, even though the IFAT was only used for part of the semester, they benefited the most. However, there is no reason to believe that any systematic WM differences exist between the classes.

Another possibility is that partial IFAT use led to the ideal amount of motivation to study outside of class, therefore increasing assignment and test grades more so than no IFAT use and full IFAT use. According to the Yerkes-Dodson law, a moderate amount of arousal leads to optimal performance, with lower and higher levels leading to impaired performance (Yerkes & Dodson, 1908). It is possible, then, that partial IFAT use best prepared students in giving them a sense of what they knew well and what material they needed to spend additional time reviewing, leading to a moderate amount of arousal and motivation to prepare in advance (i.e., prior to working on assignments and taking tests). In the no IFAT use condition, this likely led to high arousal such that students may have felt academically underprepared, leading to higher anxiety levels that interfered with information processing and cognitive performance. In the full IFAT use condition, this could have led to low arousal such that students may have felt overconfident prior to completing assignments and in preparing for tests, leading to lower motivation to study and less effort. Similarly, with full IFAT use, it is possible that students relied more heavily on IFAT used during the weekly review sessions to retain information, which decreased motivation to independently review the course material outside of class.

Related, it is possible that full IFAT use led to the Dunning-Kruger effect, in which the more one lacks skills or knowledge in a specific domain, the more likely one is to overestimate one's skill or knowledge level in that domain (Kruger & Dunning, 1999). This effect occurs because when the target lacks skill and knowledge, they are less able to differentiate between what constitutes weak versus strong performance. In comparing the partial and full IFAT use conditions, partial IFAT use likely led participants to accurately gauge their skill and knowledge level in the course content by allowing them to compare their own performance before using the IFAT (first half of the semester) and after using the IFAT (second half of the semester). This comparison, then, showed participants gaps in their knowledge and encouraged better study skills, which would predict moderate levels of arousal in motivation to study, resulting in optimal performance. Since the IFAT was used consistently and all throughout the semester in the full IFAT use condition, this did not allow for students to compare their own performance between not using and using the IFAT as a learning tool, which would be less likely to expose gaps in their knowledge or reveal to them that they had not been adequately studying in completing assignments and preparing for tests. This then would lead to the Dunning-Kruger effect, falsely inflating confidence in their knowledge level and perceptions of their skill, thus predicting lower levels of arousal leading to lower performance.

Another potential explanation is related to group dynamics, specifically social loafing. Social loafing is the tendency for individuals in group settings to exert less effort compared to individuals working alone. This effect has been documented in academics, athletics, social situations, and professional work settings (Karau & Williams, 1993). In the full IFAT condition, with repeated opportunity to engage in group discussions, students likely developed the expectation that they would be given class time to collaborate in groups for the purposes of reviewing the content and thus may have relied more heavily on group members to "teach them" the material, leading to participants exerting less effort in independently reviewing the course content. In the partial IFAT use condition, these group discussions were introduced halfway through the semester, making it more likely that students had already developed their own method of preparing for the test, and could therefore show a decreased dependence on other students to assist them in reviewing the material.

We also did not keep track of absent students or of the level of participation or engagement in the activity for each student, so it could also be the case that more absences and/or less engagement occurred in the full IFAT condition, resulting in a smaller effect. These variables were not included to protect student anonymity. Future research should tease apart these potential effects by controlling for these variables during the investigation. Additionally, future investigations would benefit from including a true experimental design with the same group of students.

It is also worth addressing that the specific tool used here (IFAT) was a paper-based one which could not be used online. However, the combination of the same learning principles (retrieval, collaboration, immediate feedback) could be achieved in an online environment by combining multiple tools. For example, assigning small groups of students (3-5 students) to breakout rooms where they can collaborate on a quiz (perhaps after the Learning Management System is set up) which provides immediate feedback and allows for partial credit for a 2nd or 3rd correct selection would likely achieve a similar environment as was created in the present activity.

Although the results here are not as clean as we had predicted, it does provide preliminary support that using the IFAT in this way can lead to improved performance over not using it. Additionally, it appears to be a valuable tool to consider using, as students do appear to enjoy using it (although it is based on informal student feedback and hearsay in this study, but has been previously reported by Kennette & McGuckin, 2018, including that students self-reported perceiving learning more when using the IFAT).

In conclusion, although we did not find the expected gains in students' performance, there are documented benefits to the types of pedagogical experiences that the IFAT allows students to engage in, such as retrieval practice, elaboration, and spaced retrieval. Further, students seemed to enjoy the activity, so it should also impact student motivation and engagement. As such, the IFAT could be included as one of many pedagogical tools at our disposal to create positive learning experiences for students.

Statements and Declarations - The author reports no conflict of interest. Ethical approval was obtained by the respective institutional review board prior to data collection.

Funding – This work was not funded by any agency or entity.

References

Adesope, O. O., Trevisan, D. A. & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659–701. https://doi.org/10.3102%2F0034654316689306

Ariel, R., & Karpicke, J. D. (2018). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied, 24*(1), 43-56. https://doi.org/10.1037/xap0000133

Baddeley, A. D., & Hitch, G. J. (2017). Is the levels of processing effect language-limited?. *Journal of Memory and Language, 92,* 1–13.

Baddeley, A. D. (1997). *Human Memory: Theory and Practice*. Psychology Press.

Bird, C.M., Keidel, J. L., Ing, L. P., Horner, A. J., Burgess, N. (2015). Consolidation of complex events via reinstatement in posterior cingulate cortex. *Journal of Neuroscience*, *35*(43), 14426-14434. https://doi.org/10.1523%2FJNEUROSCI.1774-15.2015

Bjork, R. A. (2011). On the symbiosis of remembering, forgetting, and learning. In A. S. Benjamin (Eds.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp. 1–22). Psychology Press. https://doi.org/10.4324/9780203842539

Bowman, M., Frame, D. L. & Kennette, L. N. (2013). Enhancing teaching and learning: How cognitive research can help. *Journal on Excellence in College Teaching: Brain-Based Learning (Special Issue), 24*(3), 7–28.

Butler, A. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 36*, 1118-1133. https://doi.org/10.1037/a0019902

Canadian Institutes of Health Research (2018). Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council, Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html

Chang, F, T. Y., & Li. M. N. F. (2019). A relative effectiveness assessment of an immediate feedback assessment technique employed on computers and scratch cards: A meta-analysis. *International Journal of Educational Technology, 6*(1), 12-29.

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W. & DiCarlo, S.E. (2003). Student retention of course content is improved by collaborative group testing. *Advanced Physiological Education, 27*(3), 102–108. https://doi.org/10.1152/advan.00041.2002

Craik, F. I. M. & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Davidson, N., Major, C. H., & Michaelsen, L. K. (2014). Small-group learning in higher education – cooperative, collaborative, problem-based, and team-based learning. *Journal on Excellence in College Teaching, 25*(3-4), 1-6.

Dihoff, R. E., Brosvic, G. M., Epstein, M. L. & Cook, M. J. (2004) Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *The Psychological Record, 54*(2), 207–231. https://doi.org/10.1007/BF03395471

Fink, L. D. (2004). Beyond small groups: Harnessing the extraordinary power of learning teams. In L. K. Michaelsen, A. B. Knight, & L. D. Fink, *Team-Based Learning* (pp. 3-26). Sterling, VA: Stylus Publishing.

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology, 38*(4), 281–288. https://doi.org/10.1016/j.cedpsych. 2013.06.001

Fiorella, L., & Mayer, R. E. (2014). Role of expectations and explanations in learning by teaching. *Contemporary Educational Psychology, 39*(2), 75–85. https://doi.org/10.1016/j.cedpsych.2014.01.001

Himmer, L., Schönauer, M., Heib, D. P. J., Schabus, M., & Gais, S. (2019). Rehearsal initiates systems memory consolidation, sleep makes it last. *Science Advances*, *5*(4), eaav1695. https://doi.org/10.1126/sciadv.aav1695

Hoogerheide, V., Deijkers, L., Loyens, S. M., Heijltjes, A., & van Gog, T. (2016). Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them. *Contemporary Educational Psychology, 44-45*, 95–106. https://doi.org/10.1016/j. cedpsych.2016.02.005

Karau, S. J. & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65,* 681–706. https://doi.org/10.1037/0022-3514.65.4.681

Kennette, L. N. & McGuckin, D. (2018). Using the Immediate Feedback Assessment Technique (IFAT) for non-assessments: Student perceptions and performance. *Psychology Teaching Review, 24*(2), 66-71.

Kruger, J. M. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence leads to inflated self-assessments. *Journal of Personality and Social Psychology, 77,* 1121-1134. https://doi.org/10.1037//0022-3514.77.6.1121

Laal, M., & Ghodsi, S. M. (2011). Benefits of collaborative learning. *Procedia- Social and Behavioral Sciences, 31,* 486-490. https://doi.org/10.1016/j.sbspro.2011.12.091

Leung, K. C. (2019). An updated meta-analysis on the effect of peer tutoring on tutors' achievement. *School Psychology International, 40*(2), 200-214. https://doi.org/10.1177%2F0143034318808832

Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review, 27*(4), 645-656. https://doi.org/10.1007/s10648-015-9312-8

Pashler, H., Cepeda, N. J., Wixted, J. T., Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 3-8. https://doi.org/10.1037/0278-7393.31.1.3

Rajaram, S. & Pereira-Pasarin, L. P. (2007). Collaboration can improve individual recognition memory: Evidence from immediate and delayed tests. *Psychonomic Bulletin & Review, 14*(1), 95–100. https://doi.org/10.3758/bf03194034

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%2Fj.1467-9280.2006.01693.x

Shute, V. J. (2008). Focus on formative feedback. *Review of Education, 78*(1), 153-189. https://doi.org/10.3102%2F0034654307313795

Slepkov, A. D., & Godfrey, A. T. K. (2019). Partial credit in answer-until-correct multiple-choice tests deployed in a classroom setting. *Applied Measurement in Education, 32*(2), 138-150. https://doi.org/10.1080/08957347.2019.1577249

Squire, L.R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021766. https://doi.org/10.1101/cshperspect.a021766

Weinstein, Y., Sumeracki, M., & Caviglioli, O. (2018). *Understanding how we learn: A visual guide*. Routledge.

Wickens, T. D., & Keppel, G. (2004). Design and analysis: A researcher's handbook. Pearson Prentice-Hall.

Yerkes R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18,* 459–482. https://doi.org/10.1002/cne.920180503