

Early, But Not Intermediate, Evaluative Feedback Predicts Cumulative Exam Scores in Large Lecture-Style Post-Secondary Education Classrooms

Alice S. N. Kim
York University

Sharry Shakory
Ontario Institute for Studies in Education,
University of Toronto

A critical step in promoting academic success among students involves the early identification of those in need of additional support before they fall too far behind. With this aim in mind, the present study investigated whether early and intermediate evaluative feedback on in-class quizzes were predictive of students' scores on a final cumulative exam in a third year Psychology course at a large North American university. Early feedback was operationally defined as the percentage score that students received on a quiz that took place on the third of 12 classes, whereas intermediate feedback was operationally defined as the percentage score that students received on a quiz that took place on the seventh of 12 classes. The results of a regression analysis showed that early, but not intermediate, evaluative feedback was predictive of students' scores on the final cumulative exam. The implications of the present findings include a practical, low-cost means of identifying students who could benefit most from additional academic support and resources to help enhance their achievement in a course. Moreover, the present study suggests that it is important for students to adopt effective study habits and learning strategies from the beginning of a course. Adopting these practices could help improve students' academic success, experience, and retention rates, benefiting both the university and student body.

Keywords: evaluative feedback, assessment, academic achievement, cumulative exam, knowledge retention

Large lecture-style classes are common in the context of post-secondary education, making it critical to identify how students' learning can be enhanced in this setting. Factors that play a key

role in impacting students' performance in the context of large post-secondary education classes include testing of students (Bijol, Byrne-Dugan, & Hoenig, 2015; Carrillo-de-la-Peña, & Pérez, 2012; McDaniel, Wildman, & Anderson, 2012) and frequency of such testing, where more testing has been linked to better academic performance (Johnson & Kiviniemi, 2009; Leeming, 2002; Myers & Myers, 2006). Importantly, testing followed by feedback has been shown to have a greater benefit on knowledge retention compared to testing without the provision of feedback (e.g., Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Cull, 2000; Pashler, Cepeda, Wixted, & Rohrer, 2005). The importance of providing feedback as part of the assessment process, and as a means to enhance student learning and academic performance, has been well established in the post-secondary education literature (Evans, 2013; Nicol & Macfarlane-Dick, 2006; Price, Handley, Millar, & O'Donovan, 2010). However,

Alice S. N. Kim, Teaching Commons, York University; Sharry Shakory, Department of Applied Psychology and Human Development, Ontario Institute for Studies in Education, University of Toronto.

We thank the Teaching Commons at York University, particularly Mandy Frake-Mistak, for providing guidance and creating a supportive environment for the scholarship of teaching and learning through the EduCATE program. We would also like to thank the members of our Community of Practice (Kalina Grewal, Paul Baxter, Linda Carozza, Eric Armstrong, Mandy Frake-Mistak) for the insights, support, and inspiration that they provided throughout every stage of this study.

Correspondence concerning this article should be addressed to Alice S. N. Kim, Teaching Commons, York University, Room 1046 Victor Phillip Dahdaleh (formerly TEL) Building, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada. E-mail: a_kim@yorku.ca

little work has been done to investigate whether evaluative feedback that students receive early on and midway through a course is predictive of their overall performance (e.g., grade) in the course. In the case that it is, these evaluative measures can be used to identify students who are at-risk for performing poorly in the course. These students can then be offered additional aids or academic intervention as early as possible to help elevate their trajectory in the course before they fall too far behind.

The present study investigated whether evaluative feedback (i.e., percentage scores) can be used to predict students' performance on a final cumulative exam, and whether this could be done using feedback corresponding to assessments given early on and in the middle of a course. Evaluative feedback provides information in regards to the correctness of a student's response (e.g., letter grades, percentage scores) and has been documented as the most common type of feedback that students receive (Marzano, 2000; Oosterhof, 2001; for examples see Brown, Bice, Shaw, & Shaw, 2015; Downs, 2015; Marden, Ulman, Wilson, & Velan, 2013). Descriptive or developmental feedback, on the other hand, provides a more detailed account of a student's performance, along with suggestions for improvement (Lipnevich & Smith, 2008). To be effective, feedback that is provided to students should be timely, relevant (Ramsden, 2003), and frequent (Phelps, 2012). Although the provision of developmental feedback is encouraged (Evans, 2013), it requires more work compared with the provision of evaluative feedback and is thus more likely to result in feedback that is untimely compared with evaluative feedback. For this reason, many course instructors with large classes may be more inclined to provide students with evaluative, as opposed to developmental, feedback. Consequently, further insight into the predictive utility of evaluative feedback on students' academic performance would be particularly useful in the context of large post-secondary education classes.

Past studies have found that marks corresponding to both early (Nowakowski, 2006; Winston, van der Vleuten, & Scherpbier, 2014), and midterm assessments are correlated with final grades (Connor, Franko, & Wambach, 2006; Jensen & Barron, 2014). However, assessments that take place throughout a course often contribute to students' final grades and

thus may confound any analyses that are focused on testing the role of evaluative feedback in predicting final grades (but see Nowakowski, 2006). An alternative outcome measure that can be used to assess the relation between feedback and academic achievement are students' final exam scores. Cumulative, compared with non-cumulative, exams serve as a better index of what students have learned throughout an entire course (Khanna, Brack, & Finken, 2013). Further research examining whether evaluative feedback is predictive of students performance on a final cumulative exam would therefore help shed light on whether this type of feedback can be used to predict students' academic success in a course.

To the best of our knowledge, only two studies have investigated how evaluative feedback relates to students' performance on a final cumulative exam (Azzi, Ramnanan, Smith, Dionne, and Jalali, 2015; Landrum, 2007). Both studies found that the mean scores of quizzes that students completed throughout a course were correlated with students' performance on a final cumulative exam. Moreover, Azzi et al. (2015) also found that midterm exam scores correlated with performance on the final cumulative exam. An inquiry that builds upon this work is an investigation of whether evaluative assessments, quizzes in particular, are predictive of final cumulative exam scores depending on the time point during the course (e.g., early vs. late during the course) that the assessments are held.

A difficulty that arises upon review of the relevant literature is operationally defining early, as opposed to intermediate, feedback and distinguishing between the two. For example, midterm and first exams can take place at different points within a course and are likely to vary across different courses within and across institutions. Moreover, it is not always made clear in past studies when exactly the assessments took place (e.g., Jensen & Barron, 2014; Kibble, 2011). This makes it difficult to interpret the results of such studies in relation to others, and to critically assess whether the results speak to the impact of early, intermediate, or even late evaluative feedback. As described in the following section, early and intermediate feedback were operationally defined in the present study based on the exact time point when the

relevant testing occurred in a course consisting of 12 classes.

Current Study

The purpose of the present study was to investigate whether evaluative feedback received by students early on and during the middle of a course could predict their performance on a final cumulative exam, particularly in the context of a large lecture-style, third year Psychology course taught at a large North American university by one of the authors. The present study adds to the extant literature by specifically assessing the predictive utility of evaluative feedback, corresponding to two different time points within a course (early and intermediary), on final cumulative exam scores. For students in need of additional support, early identification and intervention is ideal to prevent them from falling behind, and would help increase student retention rates as well as enhance the student experience. Thus, the results of the present study may have important implications for education, including how courses are designed and structured, as well as providing a basis for identifying students who could benefit most from additional support or academic intervention early on during a course.

In the present study, early feedback was operationally defined as the percentage score that students received on the first of five quizzes, which took place on the third of 12 classes. Intermediate feedback was operationally defined as the percentage score that students received on the third of five quizzes, which took place on the seventh of 12 classes. Based on past studies that found evaluative feedback, in general, to be positively correlated with student academic achievement, we hypothesized that both early and intermediate feedback would predict students' performance on a final cumulative exam.

Method

Participants

All students were enrolled in a section of a third year Psychology course taught at a large Canadian university by the first author. Students were invited to participate in the study after the final marks for the course were submitted to the university. All students who

participated in the study consented to have their data from the course included in the study, following the procedures approved by the Human Participants Review Sub-Committee, York University's Ethics Review Board. Out of 282, 41 students who completed the course in one of two semesters consented to participate in the study (14 students from the summer semester and 27 students from the consecutive fall semester). The same instructor taught the course in both semesters. Of the 41 participants included in the study, 31 were female, and the mean age of the students was 25 years (18 to 54 years; $SD = 8.08$ years).

Materials

In-class quizzes. Quiz questions consisted of questions selected from a question bank provided by the publisher of the textbook used in the class (Galotti, Fernandes, Fugelsang, & Stolz, 2010), modified versions of these questions, and questions that the instructor came up with based on the lectures. The quiz questions were selected by the course instructor to help prepare students, in terms of content and difficulty, for the final cumulative exam. The question formats used consisted of multiple choice and short answer. The short-answer questions required students to list, identify, and/or explain theories and key concepts covered in class. All quiz questions corresponded to the learning objectives that were assigned for the material that was covered and required students to both retrieve the correct information from memory and to apply their knowledge of the course material. The learning objectives for each class were provided to students before and at the beginning of each class. Students were given 30 min to provide their answers to the questions included on each quiz. The number of questions that the quizzes comprised depended on the number of components involved in answering the short-answer questions. However, the quizzes were created such that students would be able to comfortably answer all the questions within the allotted amount of time. Students' responses to these questions were scored using detailed answer keys provided by the course instructor. For each question, the answer key specifically indicated what students' responses had to include for both full and partial marks. The quizzes were marked by teaching assistants and, in

some cases, the course instructor. The course instructor checked a sample of the students' quizzes for consistency and accuracy of marking and found 100% agreement with the marks allotted by the teaching assistants. This high level of agreement was attributed to a combination of conscientious marking on the part of the teaching assistants and the high level of detail provided in the answer keys. The evaluative feedback provided to students after each quiz was in the form of a percentage score that was calculated using the marks that students received for providing correct responses. Students' percentage scores on the first and third quizzes were used as a measure of early and intermediary feedback, respectively.

Exam. The final exam was cumulative, spanning all of the content covered throughout the course and consisting of multiple choice and short-answer questions. Similar to the quizzes, the exam questions required students to both retrieve the correct information from memory and to apply their knowledge of the course material. Additionally, the questions corresponded to the learning objectives provided to students and consisted of questions that were selected from a question bank provided by the publisher of the textbook used in the class (Galotti et al., 2010), modified versions of these questions, and questions that the instructor came up with based on the lectures. Similar to the quizzes, the short-answer questions required students to list, identify, and/or explain theories and key concepts covered in class, and their responses were scored using a detailed answer key provided by the course instructor. The exams were marked by teaching assistants and, in some cases, the course instructor. Like the quizzes, the course instructor checked a sample of the students' exams for consistency and accuracy of marking. Students had three hours to complete the exam, and an overall exam score percentage was calculated for each student based on the marks they earned for providing correct answers on the exam.

Procedure

For both semesters, a 3-hr class period was used to cover a chapter of the textbook and other relevant material. The summer semester consisted of two 3-hr class periods per week, whereas the fall semester consisted of one 3-hr

class period per week. For both semesters, the classes were held at the same time in the evening. The same chapters of the same textbook (Galotti et al., 2010) were covered across both semesters, and the evaluation breakdown was also the same for both semesters. Students in both the summer and fall sections of the course completed five quizzes that each covered two chapters of the textbook and the corresponding lectures. A quiz was held every second class, for a total of five quizzes. Thus, students in the summer semester completed a quiz each week, whereas students in the fall semester completed a quiz every second week. Students in both sections were notified at the beginning of the course, through class announcements and the course syllabus, that their best three out of five quizzes would be used to calculate 50% of their final mark and that the final cumulative exam would be worth 40% of their final mark. Students in both sections completed a final cumulative exam during the designated exam period, and the same instructor taught both sections of the course.

Multiple Regression Analysis

A regression analysis was conducted to assess whether early and intermediary feedback (scores on the first and third quizzes, respectively) were predictive of students' performance on the final cumulative exam. Since the data included in the analyses were collected across two (summer and fall) sections of the course in different semesters, semester was also included as a factor in the regression analysis.

Results

Table 1 present the means, standard deviations, maximum and minimum scores, and range for the first and third quizzes, as well as the final cumulative exam. The ranges for each of these variables are fairly large, which is the typical distribution for a large course at this university.

Regression Model

A multiple linear regression was calculated to predict students' scores on a cumulative final exam based on the following three factors: (a) scores on the first quiz (QUIZ 1); (b) scores on

Table 1
Descriptive Data for Quiz 1 (Early Feedback), Quiz 3 (Intermediary Feedback), and the Final Cumulative Exam as Percentages

Factor	<i>M</i>	<i>SD</i>	Min	Max	Range
Quiz 1	75.51	15.03	34.4	97.8	63.4
Quiz 3	78.61	13.79	46.7	97.8	51.1
Final exam	76.72	14.21	47.8	98.7	50.9

Note. *M* = mean score; *SD* = standard deviation; Min = minimum score; Max = maximum score.

the third quiz (QUIZ 3); (c) the semester in which the course was taken (summer vs. fall). An analysis of standard residuals revealed that the data did not contain any outliers (Std. residual min = -1.862 , Std. residual max = 1.629). When the assumption of collinearity was tested, the results demonstrated that multicollinearity was not a concern (Quiz 1, tolerance = $.993$, *VIF* = 1.007 ; Quiz 3, tolerance = $.984$, *VIF* = 1.016 ; class, tolerance = $.990$, *VIF* = 1.010). The data also met the assumption of independent errors (Durbin-Watson value = 1.264). The 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 nonzero variances (Quiz 1, variance = 226.043 ; Quiz 3, variance = 844.713 ; class, variance = $.230$; final exam, variance = 201.834).

Using the enter method, a significant regression equation was found, $F(3, 37) = 5.513$, $p = .003$, with an R^2 of $.309$. Participants' predicted scores on the final cumulative exam is equal to $32.216 + .501(\text{QUIZ 1 SCORE}) + .060(\text{QUIZ 3 SCORE}) + 1.538(\text{SEMESTER})$, where scores on the first and third quizzes are measures in percentage, and semester is coded as 1 = summer, 2 = fall. Participants' predicted cumulative exam scores increased $.501\%$ for each percentage point of the first quiz, $\beta = .530$, $t(37) = 3.864$, $p < .001$, $sr^2 = .279$, and by $.060\%$ for each percentage point of the third quiz, $\beta = .123$, $t(37) = .890$, $p = .379$, $sr^2 = .015$. Students in the summer class scored 1.538 percentage points lower than students in the fall class, $\beta = .052$, $t(37) = .379$, $p =$

$.707$, $sr^2 = .003$. However, only the scores on the first quiz were found to be a significant predictor of students' performance on the final cumulative exam.

Discussion

The present study demonstrates that evaluative feedback provided early on during a course of study can help predict students' performance on a final cumulative exam. The implications of this finding include a practical, low-cost means of identifying students who could benefit most from additional academic support and resources to help enhance their achievement in a course. The present study is the first, to our knowledge, to investigate the predictive utility of evaluative feedback on students' performance on a subsequent final cumulative exam, and whether the predictive nature of evaluative feedback is dependent on when the corresponding assessment took place during a course of study (i.e., early vs. intermediary feedback). We relate the findings of the present study to the results of relevant past studies in the discussion below. We then discuss possible explanations for the present findings, as well as potential implications for the promotion of students' academic success, retention rates, and a positive student experience in the context of post-secondary education.

The results of the present study are in line with those of past studies demonstrating that early marks are correlated with final grades. For example, [Winston and colleagues \(2014\)](#) found that an exam administered in the first two weeks of a course predicted whether medical school students passed or failed the course, and [Nowakowski \(2006\)](#) found that early assessment grades were highly correlated with final grades. However, past studies have also shown that

midterm marks are correlated with final grades (Connor et al., 2006; Jensen & Barron, 2014), which does not align with our finding that intermediate feedback does not predict performance on a final cumulative exam. This inconsistency could be due to differences in the outcome measure used. Whereas a final cumulative exam assesses students' knowledge of the course material in one instance, final grades are typically composed of assessments corresponding to different segments of the course that took place across multiple occasions. However, Azzi and colleagues (2015) also found a significant correlation between students' scores on a midterm and final cumulative exam, which, again, does not align with the results of the present study.

The study by Azzi and colleagues (2015) was conducted in the context of a modified team based anatomy laboratory for a medical school, and the midterm test format used was a "bell-ringer." During a "bell-ringer" examination, students move from one station to the next and are given a specified amount of time to answer questions at each station. In addition to these aspects of the course design, another major factor distinguishing this study from the present study was that the midterm was worth 30% of students' final mark for the course, whereas the intermediate feedback that students received in the present study corresponded to a quiz that could potentially be counted as one of their best three out of five quizzes that would account for 50% of their final mark. Thus, the results of the present study may differ from that of Azzi and colleagues due to any number and combination of factors, including the nature of the course (lecture style vs. modified team based learning), test format, and the weight of the relevant assessment on students' final marks. Future research should investigate further the potential impact of each of these factors on the predictive utility of evaluative feedback on students' summative learning outcomes.

A possible explanation for the present findings could be related to the course design of the present study, specifically that the best three out of five quizzes were used to calculate students' final grades. Specifically, students may have used different strategies in regards to how they took the course. For example, if a student scored well on the first two quizzes and they had exams or assignments due for other courses around the

same week as the third quiz, they may have opted to focus their available time on the assignments or exams for the other courses, knowing that they could drop the third quiz. In contrast, it could be that students became less engaged with a course over time. Students often try their best during the beginning of a course, which is also a time when their workload is typically not as heavy compared with the middle of a course or semester. Thus, it may be that students who were engaged with the course early on earned good grades, and then later, after a period of flagged engagement, these same students prepared effectively for the cumulative exam, resulting in good exam performance that correlated with their good performance from earlier on in the course. On the other hand, it could be that students who were not engaged with the course from the beginning remained disengaged throughout and did not effectively prepare for the exam. This could have then resulted in poor performance on the exam as well as the first and third quizzes. In this scenario, students' marks for the first quiz and the final exam would have likely correlated regardless of whether the students were engaged at the beginning of the course followed by a period of flagged engagement or disengaged from the course from the very beginning.

Alternatively, one might attribute the lack of significance found for the predictive utility of intermediate feedback on cumulative exam scores to a lack of power. Indeed, it is possible that increasing our sample size may change our results. An a priori power analysis revealed that a sample consisting of 36 participants would be required to detect a large effect ($f^2 = .35$) with 80% power. Alternatively, to detect medium and small effects ($f^2 = .15$ and $.02$, respectively) with 80% power, a sample consisting of 77 and 550 participants, respectively, would be required. Based on these parameters, it is possible that we were not able to find a significant relation between intermediate feedback and the final cumulative exam due to a lack of power for medium and small effects. If, however, there was a large effect for intermediate feedback, our sample of 41 participants should have been sufficient to pick up on this finding. Future research will have to be conducted to investigate this issue further. The remaining discussion will focus on the finding that early feedback was

found to be predictive of students' performance on the final cumulative exam.

In terms of applying the main finding of the present study in the context of post-secondary education, course instructors should be made aware of the benefits of implementing assessments accompanied with evaluative feedback to students early, within the first few classes, during a course of study, and the potential use of assessments as a tool for identifying students who may require additional support. Conversely, it is also informative for students to be made aware of the predictive relation between early evaluative feedback and performance on a final cumulative exam, as this would help emphasize the importance of doing well from the beginning of the course. Moreover, students can use this information to assess whether their current study method is sufficient for the goals they have for the course in question or whether they need to reassess and modify their study methods. If students are dissatisfied with the marks they earned early on in the course, they should take this as a cue to adopt different, more effective, study methods to change their trajectory in the course and increase their probability of achieving more favorable results on the final exam. Low-stakes, early evaluative feedback could be particularly useful if used in this manner.

In addition to being made aware of the importance of adopting effective study habits and learning strategies from the start of a course, all students would likely benefit from workshops and other forms of instruction that outline the fundamentals of effective study habits and learning strategies, especially if taken upon entry into post-secondary education, along with refresher courses or reminders at each level of study. For example, students would likely benefit from being explicitly told that among the various learning strategies, the surface learning strategy, which entails rote or shallow encoding of course material, is ineffective compared with deep and meaningful encoding strategies for subsequent memory of the relevant information (Craig & Lockhart, 1972) and assessment-based academic achievement (Bickerdike, O'Deasmhunaigh, O'Flynn, & O'Tuathaigh, 2016; Clarke & McKenzie, 1994). Moreover, time management has been shown to be particularly important for establishing good study habits in post-secondary education (Bickerdike et al., 2016; West & Sadoski, 2011). Future research should

investigate the efficacy of interventions that target self-regulated learning skills and time management in improving academic achievement among students in post-secondary education.

Students' intrinsic motivation also plays an important role in their learning and academic success (e.g., Spinath, Spinath, Harlaar, & Plomin, 2006; Steinmayr & Spinath, 2009). Negative feedback on a test may lower the motivation levels of students who perceive themselves to have low ability (vs. students who perceive themselves as having high ability; Weidinger, Spinath, & Steinmayr, 2016). For this reason, course instructors should be encouraged to incorporate instructional strategies that promote students' motivation in a given course. This may include strategies that promote the following perceptions among students: students have control over their own learning, the course content is useful, and that effort will be rewarded with success (Jones, 2009).

In the present study, many students enrolled in the course (not necessarily those who participated in the study) communicated to the course instructor that they found the frequent testing, in the form of quizzes, helpful because it motivated them to keep up with the course readings and lecture materials. Along these lines, Landrum (2007) found improved performance when students' mean quiz scores were compared with their final cumulative exam scores, and that students who fell within the bottom third of the class with regard to their overall quiz performance made the most gains on the final exam. Past studies have also explored student perspectives on the assessment and/or feedback that they received during a course of study, and have typically found that the majority of students report such processes to positively contribute to their academic engagement, learning, and/or motivation (Bälter, Enström, & Klingenberg, 2013; Carrillo-de-la-Peña & Pérez, 2012; Healy, McCutcheon, & Doran, 2014; Leeming, 2002; Marriott & Lau, 2008). Interestingly, when it comes to views on the *purpose* of assessment, differences in the extent to which assessments are perceived to be relevant over long periods have been noted. Many students perceive the long-term relevance of assessments as a means of developing one's knowledge base and understanding of subject matter, whereas others tend to perceive only a short-term relevance for the role of assessment

(e.g., to provide grades). This again demonstrates the importance of emphasizing to all students that assessment is an important component of the learning process and can be used as a tool to achieve learning goals and master effective study habits (Healy et al., 2014).

As mentioned above, future studies will have to be conducted to investigate further the predictive utility of intermediate feedback on cumulative exam scores. These studies should assess the impact of the nature of the course, test format, and weight of the relevant assessment on students' final marks, in addition to ensuring a sufficient sample size to pick up even small and medium effect sizes. A limitation of the present study is the potentially biased nature of the sample. Students were asked to participate in the study after the final marks for the course had been submitted to the university. Thus, it could be that students who enjoyed and/or did well in the course were more likely to participate in the study, which could have resulted in skewed data. The sample was also predominantly made up of female participants, which could have also skewed the data. For example, past research has shown that females benefit more than males from completing biweekly quizzes as opposed to a midterm exam (Myers & Myers, 2006). Because of the large proportion of females in the present sample, the results of the present study may be more generalizable to females compared with males. Future studies will have to look into whether there is any impact of having a larger proportion of males in the sample.

Conclusion

The present study demonstrated that early, but not intermediate, evaluative feedback is predictive of students' performance on a final cumulative exam. Early feedback was operationally defined as the percentage score that students received on a quiz that took place on the third of 12 classes. Intermediate feedback was operationally defined as the percentage score that students received on a quiz that took place on the seventh of 12 classes. The predictive quality of early evaluative feedback may have reflected students' motivation, study habits, and learning strategies that they began the course with. The implications of the present findings include a practical, low-cost means of

identifying students who could benefit most from additional academic support and resources to help enhance their achievement in a course. Moreover, the results of the present study highlight the importance for students to adopt effective study habits and learning strategies from the beginning of a course. The outcome of such actions would likely help to improve the student experience and retention rates, benefiting both the university and student body.

References

- Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III, & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology*, *22*, 861–876. <http://dx.doi.org/10.1002/acp.1391>
- Azzi, A. J., Ramnanan, C. J., Smith, J., Dionne, É., & Jalali, A. (2015). To quiz or not to quiz: Formative tests help detect students at risk of failing the clinical anatomy course. *Anatomical Sciences Education*, *8*, 413–420. <http://dx.doi.org/10.1002/ase.1488>
- Bälter, O., Enström, E., & Klingenberg, B. (2013). The effect of short formative diagnostic web quizzes with minimal feedback. *Computers & Education*, *60*, 234–242. <http://dx.doi.org/10.1016/j.compedu.2012.08.014>
- Bickerdike, A., O'Deasmhunaigh, C., O'Flynn, S., & O'Tuathaigh, C. (2016). Learning strategies, study habits and social networking activity of undergraduate medical students. *International Journal of Medical Education*, *7*, 230–236. <http://dx.doi.org/10.5116/ijme.576f.d074>
- Bijol, V., Byrne-Dugan, C. J., & Hoenig, M. P. (2015). Medical student web-based formative assessment tool for renal pathology. *Medical Education Online*, *20*, 26765–26772. <http://dx.doi.org/10.3402/meo.v20.26765>
- Brown, G. A., Bice, M. R., Shaw, B. S., & Shaw, I. (2015). Online quizzes promote inconsistent improvements on in-class test performance in introductory anatomy and physiology. *Advances in Physiology Education*, *39*, 63–66. <http://dx.doi.org/10.1152/advan.00064.2014>
- Carrillo-de-la-Peña, M. T., & Pérez, J. (2012). Continuous assessment improved academic achievement and satisfaction of Psychology students in Spain. *Teaching of Psychology*, *39*, 45–47. <http://dx.doi.org/10.1177/0098628311430312>
- Clarke, D. M., & McKenzie, D. P. (1994). Learning approaches as a predictor of examination results in pre-clinical medical students. *Medical Teacher*, *16*, 221–227. <http://dx.doi.org/10.3109/01421599409006734>

- Connor, J., Franko, J., & Wambach, C. (2006). *A brief report: The relationship between mid-semester grades and final grades*. Minneapolis, MN: University of Minnesota.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671–684. [http://dx.doi.org/10.1016/S0022-5371\(72\)80001-X](http://dx.doi.org/10.1016/S0022-5371(72)80001-X)
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*, 215–235. [http://dx.doi.org/10.1002/\(SICI\)1099-0720\(200005/06\)14:3<215::AID-ACP640>3.0.CO;2-1](http://dx.doi.org/10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1)
- Downs, S. D. (2015). Testing in the college classroom: Do testing and feedback influence grades throughout an entire semester? *Scholarship of Teaching and Learning in Psychology*, *1*, 172–181. <http://dx.doi.org/10.1037/std0000025>
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, *83*, 70–120. <http://dx.doi.org/10.3102/0034654312474350>
- Galotti, K. M., Fernandes, N., Fugelsang, J., & Stolz, J. (2010). *Cognitive psychology: In and out of the laboratory* (First Canadian Edition). Toronto, Ontario, Canada: Nelson Education.
- Healy, M., McCutcheon, M., & Doran, J. (2014). Student views on assessment activities: Perspectives from their experience on an undergraduate programme. *Accounting Education*, *23*, 467–482. <http://dx.doi.org/10.1080/09639284.2014.949802>
- Jensen, P. A., & Barron, J. N. (2014). Midterm and first-exam grades predict final grades in biology courses. *Journal of College Science Teaching*, *44*, 82–89. http://dx.doi.org/10.2505/4/jcst14_044_02_82
- Johnson, B. C., & Kiviniemi, M. T. (2009). The effect of online chap. quizzes on exam performance in an undergraduate social psychology course. *Teaching of Psychology*, *36*, 33–37. <http://dx.doi.org/10.1080/00986280802528972>
- Jones, B. D. (2009). Motivating students to engage in learning: The MUSIC model of academic motivation. *International Journal on Teaching and Learning in Higher Education*, *21*, 272–285.
- Khanna, M. M., Brack, A. S. B., & Finken, L. L. (2013). Short- and long-term effects of cumulative finals on student learning. *Teaching of Psychology*, *40*, 175–182. <http://dx.doi.org/10.1177/0098628313487458>
- Kibble, J. D. (2011). Voluntary participation in online formative quizzes is a sensitive predictor of student success. *Advances in Physiology Education*, *35*, 95–96. <http://dx.doi.org/10.1152/advan.00053.2010>
- Landrum, R. E. (2007). Introductory psychology student performance: Weekly quizzes followed by a cumulative final exam. *Teaching of Psychology*, *34*, 177–180. <http://dx.doi.org/10.1080/00986280701498566>
- Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology*, *29*, 210–212. http://dx.doi.org/10.1207/S15328023TOP2903_06
- Lipnevich, A. A., & Smith, J. K. (2008). *Response to assessment feedback: The effects of grades, praise, and source of information*. <http://dx.doi.org/10.1002/j.2333-8504.2008.tb02116.x>
- Marden, N. Y., Ulman, L. G., Wilson, F. S., & Velan, G. M. (2013). Online feedback assessments in physiology: Effects on students' learning experiences and outcomes. *Advances in Physiology Education*, *37*, 192–200. <http://dx.doi.org/10.1152/advan.00092.2012>
- Marriott, P., & Lau, A. (2008). The use of on-line summative assessment in an undergraduate financial accounting course. *Journal of Accounting Education*, *26*, 73–90. <http://dx.doi.org/10.1016/j.jaccedu.2008.02.001>
- Marzano, R. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum and Development.
- McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition*, *1*, 18–26. <http://dx.doi.org/10.1016/j.jarmac.2011.10.001>
- Myers, C. B., & Myers, S. M. (2006). Assessing assessment: The effects of two exam formats on course achievement and evaluation. *Innovative Higher Education*, *31*, 227–236. <http://dx.doi.org/10.1007/s10755-006-9020-x>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*, 199–218. <http://dx.doi.org/10.1080/03075070600572090>
- Nowakowski, J. (2006). An evaluation of the relationship between early assessment grades and final grades. *College Student Journal*, *40*, 557–561.
- Oosterhof, A. (2001). *Classroom applications of educational measurement*. Upper Saddle River, NJ: Merrill Prentice Hall.
- 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*, 3–8. <http://dx.doi.org/10.1037/0278-7393.31.1.3>
- Phelps, R. P. (2012). The effect of testing on student achievement, 1910–2010. *International Journal of Testing*, *12*, 21–43. <http://dx.doi.org/10.1080/15305058.2011.602920>

- Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort but what is the effect? *Assessment & Evaluation in Higher Education*, *35*, 277–289. <http://dx.doi.org/10.1080/02602930903541007>
- Ramsden, P. (2003). *Learning to teach in higher education*. London, England: RoutledgeFalmer.
- Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, *34*, 363–374. <http://dx.doi.org/10.1016/j.intell.2005.11.004>
- Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, *19*, 80–90. <http://dx.doi.org/10.1016/j.lindif.2008.05.004>
- Weidinger, A. F., Spinath, B., & Steinmayr, R. (2016). Why does intrinsic motivation decline following negative feedback? The mediating role of ability self-concept and its moderation by goal orientations. *Learning and Individual Differences*, *47*, 117–128. <http://dx.doi.org/10.1016/j.lindif.2016.01.003>
- West, C., & Sadoski, M. (2011). Do study strategies predict academic performance in medical school? *Medical Education*, *45*, 696–703. <http://dx.doi.org/10.1111/j.1365-2923.2011.03929.x>
- Winston, K. A., van der Vleuten, C. P. M., & Scherpbier, A. J. J. A. (2014). Prediction and prevention of failure: An early intervention to assist at-risk medical students. *Medical Teacher*, *36*, 25–31. <http://dx.doi.org/10.3109/0142159X.2013.836270>

Received October 5, 2016

Revision received May 9, 2017

Accepted May 15, 2017 ■