

# Multiple-Choice Testing Using Immediate Feedback—Assessment Technique (IF AT<sup>®</sup>) Forms: Second-Chance Guessing vs. Second-Chance Learning?

Jeremy D. Merrel<sup>1</sup>, Pier F. Cirillo<sup>2</sup>, Pauline M. Schwartz<sup>2</sup> & Jeffrey A. Webb<sup>3</sup>

<sup>1</sup> Department of Psychology, Southern Connecticut State University, New Haven, USA

<sup>2</sup> Department of Chemistry & Chemical Engineering, University of New Haven, West Haven, USA

<sup>3</sup> Department of Chemistry, Southern Connecticut State University, New Haven, USA

Correspondence: Jeffrey A. Webb, Department of Chemistry, Southern Connecticut State University, 501 Crescent Street, New Haven, CT 06515, USA. Tel: 1-203-392-5018. E-mail: webbj6@southernct.edu

Received: July 16, 2015

Accepted: August 10, 2015

Online Published: September 17, 2015

doi:10.5539/hes.v5n5p50

URL: <http://dx.doi.org/10.5539/hes.v5n5p50>

## Abstract

Multiple choice testing is a common but often ineffective method for evaluating learning. A newer approach, however, using Immediate Feedback Assessment Technique (IF AT<sup>®</sup>, Epstein Educational Enterprise, Inc.) forms, offers several advantages. In particular, a student learns immediately if his or her answer is correct and, in the case of an incorrect answer, has an opportunity to provide a second response and receive partial credit for a correct second attempt. For a multiple choice question with five possible answers, the IF AT<sup>®</sup> form covers spaces labeled A through E with a thin opaque film; when the film is scratched away, a star indicates the correct answer. This study was conducted in order to assess learning after an initial incorrect answer. Based on random chance, students should have mathematically a 25% chance of guessing a correct second answer (i.e. 1 of 4 remaining answers on the IF AT<sup>®</sup> form). Analysis of second responses for 8775 questions on IF AT<sup>®</sup> forms in 22 classes over 3 years showed that the percent of correct second answers was 44.9%, significantly higher than one might expect from random guessing. This indicates that students learned from an incorrect answer and, possibly by re-reading the problem, were able to demonstrate some level of mastery of the material. This data leads us to conclude that IF AT<sup>®</sup> forms are useful assessment tools.

**Keywords:** exams, second-chance, partial credit, feedback, assessment

## 1. Introduction

Multiple-choice exams are advantageous with large class sizes, the desire to test frequently, and the desire to return corrected exams quickly to students. Multiple-choice exams, however, have several disadvantages for testing mastery of course material including having only one opportunity to respond to a question. The Immediate Feedback Assessment Technique (IF AT<sup>®</sup>) form, introduced by M. Epstein, B. Epstein and Brosvic (2001), attempted to circumvent these issues. When using the form, a student scratches off an opaque film corresponding to the answer to a multiple choice question and observes either a star for a correct answer or a blank for an incorrect answer; instructors have the codes for different IF AT<sup>®</sup> forms so as to write multiple choice questions where the correct answers correspond to the position of the stars on the IF AT<sup>®</sup> form. Thus, with the IF AT<sup>®</sup> form, a student receives an immediate response for correct answers; well-prepared students are particularly encouraged by this procedure (Dibattista & Gosse, 2006). But, if the area under the patch is blank (an incorrect answer), the student can immediately reconsider the question and give a second-chance answer, for which some partial credit may be awarded. More information about IF AT<sup>®</sup> forms can be found at the Epstein Educational Enterprise, Inc. web pages (<http://www.epsteineducation.com/home/>), and in an excellent overview by Smith (2013).

Nicol and Macfarlane-Dick (2006) explore the importance of learning through assessment and explain the need for a shift in focus from simply assessing a student's knowledge to a focus on a lifetime of learning. Fischer (1999) advocates that lifelong learning is fostered best utilizing self-directed learning environments. Fisher (1999) further explains that this self-directed learning should utilize authentic, complex problems and be embedded in a rewarding endeavor. For students to become self-directed learners they must monitor and adjust

their approaches to learning (Ambose, Bridges, DiPietro, Lovett, & Norman, 2010). The Immediate Feedback Assessment Technique allows for students to monitor and adjust their approach as they solve a problem (if they get their first answer wrong) and do so under the umbrella of a rewarding endeavor. Although students benefit by knowing immediately whether a mistake has been made in answering a question and by having the opportunity to receive partial credit for a correct second answer (Epstein et al., 2002; Dihoff, Brosvic, Epstein, & Cook, 2004), it is less clear if a correct second answer is indeed the result of learning from a mistake or good guessing. Which leads to our primary research problems: How accurately can students respond to a question in which they have already responded incorrectly? Can we ascertain whether students, when given a second-chance, are merely guessing or are utilizing the learning triggered by the knowledge of an incorrect answer? In a problem-solving situation, immediate knowledge of an incorrect response to a problem should trigger student's reappraisal of the information explicitly and implicitly embedded in the question. This ideally should result in a cognitive reevaluation of the question, assisting the student to find a more plausible answer to the same question on the second attempt. Whether the credits earned through the second attempts are attributed to learning from mistakes or random guessing needs an answer. Starting with the assumption that informed second chance responses would gain more partial credit points than random-chance-guess responses, we attempted to empirically examine whether the students additional credits earned in the present study were significantly greater than the points students might have earned from utilizing blind-guessing for the second chance responses. This statistical study was taken on to assess learning after an initial incorrect answer using the IF AT<sup>®</sup> format.

## 2. Methods

IF AT<sup>®</sup> forms were purchased from Epstein Educational Enterprise, Inc. (Cincinnati, OH, <http://www.epsteineducation.com/home/>). Of several types of forms, this study used forms with 5 possible answers. Different versions of the forms come with instructor-only answer keys to construct tests with correct answers in the correct positions. New forms have come with opaque backing to insure even greater resistance to any possible "see through".

The study was conducted over a 3-year period in second-year Organic Chemistry courses. A total of 22 classes used the IF AT<sup>®</sup> forms for three hourly exams and a final given to a total of 1,449 students. The multiple-choice portion of each test consisted of between 15 and 25 questions with 5-part answers. There were a total of 26,175 questions and 8,775 questions were incorrectly answered on the first attempt; the latter were evaluated for correct second attempts.

P-Values were obtained in order to determine the percentage of students that utilized the opportunity for partial credit at a greater than random chance frequency. These p-values were obtained for each individual test as well as utilizing the combination of all tests into a large pool. Each p-value was obtained by taking the difference of the normal distribution and the value of 1.0.

Further analysis was conducted through the use of a *t*-test in order to determine whether there was any significance in the difference between the total amount of partial credit actually earned throughout the study versus the total amount of partial credit that might have been earned had students only earned partial credit through random chance. Finally, a Pearson's Correlation was conducted to show the strength between the two variables of a student who scored with partial credit and the amount of partial credit that student earned.

## 3. Results

Using the 5-answer IF AT<sup>®</sup> forms, the likelihood of a student getting a question correct on the second try by random guessing would be 25% (1 correct answer/4 remaining choices). In the current study with 8,775 such questions, one could determine that random guessing would give a correct answer 2194 times (See Figure 1). In reality, students answered correctly on a second attempt a total of 3938 times or 44.9% of the time. While it is not possible to rule out that guessing had any effect on this number, our data suggests that students utilized some technique and prior knowledge to decipher the correct answers. It is important however to note that our data is only indicative of learning and not a confirmation that second-chance correct responses are learning-driven.

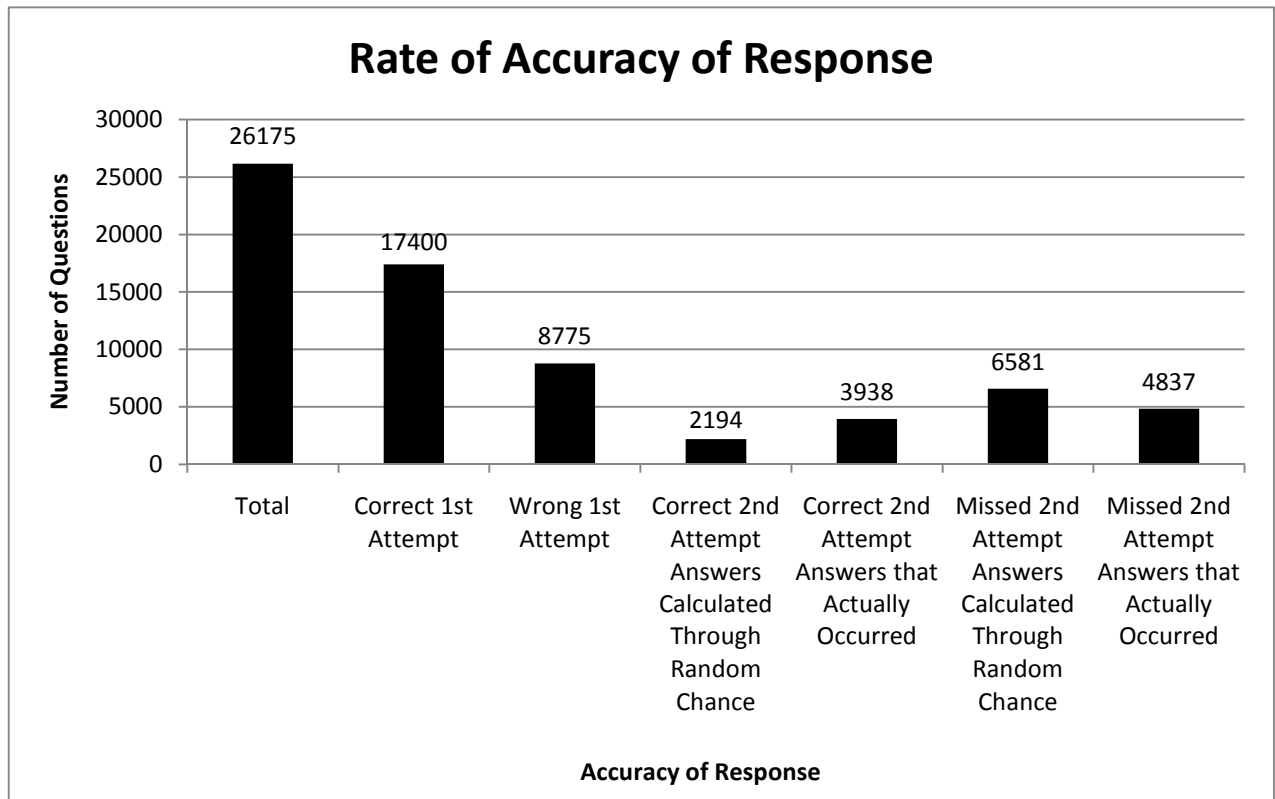


Figure 1. Rate of accurate of response actually obtained versus random chance

A *t*-test was run to determine the level of difference between the total partial credit actually earned (N=8775, Mean=179.00, SD=46.18, CI=1.27) and the value for 25% of the total possible credit that would hypothetically be obtained through random guessing (N=8775, Mean=99.72, SD=20.53, CI=0.56). The *t*-test returned a value of  $9.08 \times 10^{-12}$  (N=8775, SD=53.44, CI=1.47). This value suggests that students were utilizing some form of discernment to determine which of their remaining possibilities was the correct choice. Of the 22 tests, the greatest value for *p*-hat obtained was  $4.117 \times 10^{-8}$ . For this specific test (referred to as exam F), there were 15 multiple-choice questions; students were given the opportunity to earn 3 points for a correct first answer, 1 point for a correct second answer, or zero points for two incorrect attempts. That would result in a total possible score of 45 points. Exam F had an average score without partial credit of 21.34 and a median score without partial credit of 21. When partial credit was included for Exam F the average score was 24.30 and the median score was 25.00. The close proximity between the average and the median suggests that there are no dramatic outliers affecting the calculations. The average test grade without the addition of any partial credit was 47.42%. The average test grade after the addition of partial credit was 53.99%. This alone brought the average grade up by nearly 6.57%, which is over half of a letter grade for the multiple-choice portion of the exam (See Figure 2 for a comparison of average test scores without partial credit, with partial credit obtained at random chance, and with the partial credit that was actually obtained for each exam).

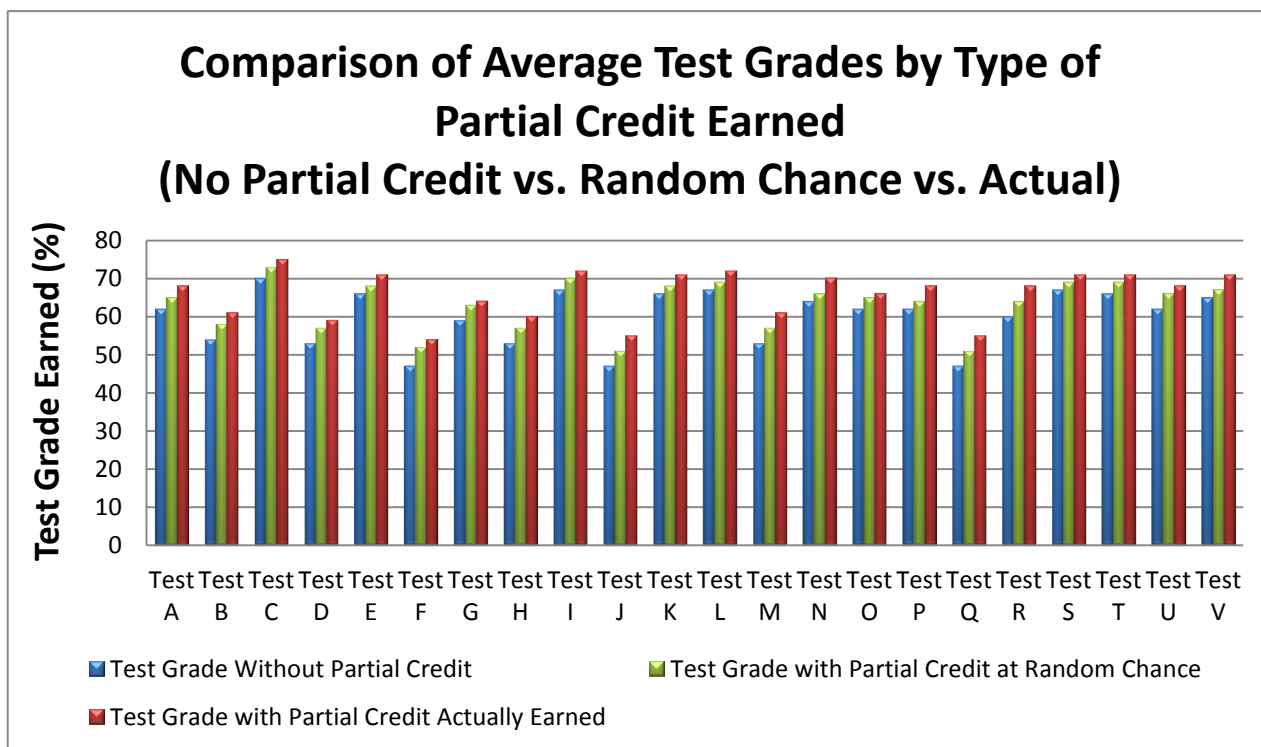


Figure 2. Comparison of average test percentage

A *t*-test was used to determine the level of significance between the scores students earned with the partial credit they actually obtained (N=44, Mean=24.30, SD=7.75, CI=3.15) versus the scores they would have theoretically earned had they only received partial credit at a random chance rate of 25% (N=44, Mean=23.31, SD=7.51, CI=3.05). A significance level of 0.01 was used to determine if there was a significant difference between the scores students earned with their actual partial credit and the scores student would have earned under random chance. The *t*-test for exam F (which, while still being significant, was the least significant of the 22 tests) displayed a value of <0.001 (N=44, SD=7.60, CI=3.09). This value strongly suggests that students were able to obtain partial credit and earn a higher grade at a statistically significant level.

The results for the Pearson’s Correlation was -0.1703, meaning that there is a weak negative correlation between the grade a student earned and the amount of partial credit that was obtained by that student. This was expected due to the understanding that to get a perfect score a student would have to earn zero partial credit and get every answer correct. Therefore, the higher grades that could be obtained would require a lower amount of incorrect initial choices.

Looking at our complete data set, the amount of partial credit points that would have been earned by guessing was 2194. In reality, students earned a total of 3938 partial credit points, resulting in an additional earning of nearly 80% more partial credit points than mere random guessing.

An analysis was completed between the average exam scores without partial credit, with partial credit obtained at random chance, and with the partial credit that was actually obtained for each exam (See Figure 2 for a visual comparison of this analysis). This analysis shows that the average grades obtained in relation to no partial credit, partial credit at random chance, and the actual partial credit earned for exams F and Q are extremely close to one another, with exam F displaying 21.34, 24.30 and 23.31 respectively; while exam Q received average exam grades of 21.17, 24.58, and 23.16 respectively. While these differences are small, and while both exams had partial credit use in statistically significant excess when compared to that of random chance, a *t*-test for exam Q resulted in a significance of  $2.3 \times 10^{-11}$  (N=52, SD=6.22, CI=2.31).

**4. Discussion and Conclusion**

The goals of teaching are to facilitate learning, to increase retention of knowledge and to expand understanding. Ideally, testing should be designed to assess learning. In today’s pedagogy, established methods of lecturing and

testing are being challenged with new approaches aimed at increasing students' understanding of course material. Methods that permit immediate feedback to students during lectures and tests have been shown to increase more effective long-term understanding (Roediger & Butler, 2011). Classroom response systems, for example, have gained considerable acceptance in engaging students during lectures in large classes (Schell, Lukoff, & Mazur, 2013; Heaslip, Donovan, & Cullen, 2014). In addition, continual "retrieval practice" enhanced with rapid feedback has been shown to assist persistent learning (Roediger & Butler, 2011). Clearly, rapid feedback is a critical element of all testing to ensure that students' mistakes do not persist (Attali, 2011).

Designing good tests is a challenge for instructors. With increasing class sizes or the desire to present multiple exam opportunities, multiple-choice questions are often used to speed up grading. Little, E. Bjork, R. Bjork and Angello (2012) note that multiple-choice tests can be useful learning tools that foster productive retrieval learning. Mathematical analysis of multiple choice tests shows that the fairest method for grading is to give credit for the number of correct answers with no penalty for a wrong or missing answer (Scharf & Baldwin, 2007); this approach, however, does not reduce credit given for blindly guessing an answer (Bush, 2015). In most cases, multiple choice questions pose two critical problems: one, students must choose from a fixed set of answers without displaying their learning process; and, two, common to all testing methods, there is a delay in learning whether the answer is correct or not.

Epstein et al. designed IF AT<sup>®</sup> forms to overcome some of these issues (Epstein et al., 2001; Epstein et al., 2002; Dihoff, Brosvic, & Epstein, 2003; Brosvic, Epstein, Cook, & Dihoff, 2005). For the professor, the immediate feedback assessment technique offers a benefit in that the grading of the multiple-choice portion of an exam is nearly completed during the actual testing process. The grading of these exams consists of a student writing the amount of points they received for each question to the side of the question after they complete the exam. In addition, many researchers have found that, when compared to other testing methods (traditional format, end-of-test feedback, delayed feedback (feedback given 24 hours after testing), and Scantron testing formats), the IF AT<sup>®</sup> provided better recall of material when asked again on a future, final exam (Epstein & Brosvic, 2002; Dihoff et al., 2003; Dihoff et al., 2004).

IF AT<sup>®</sup> forms can be used to give partial credit for a correct answer after an initial incorrect response, and some faculty use IF AT<sup>®</sup> forms to allow students to answer until he/she reaches a correct answer (Attali, 2011). One problem with giving partial credit is promoting guessing. Slepkov (2013) suggested that second answers on IF AT<sup>®</sup> forms were better than guessing. Our analysis suggests that Slepkov's assertion is valid. We found that the percent of correct second answers was 44.9%, almost 25% higher than random guessing. With rare exceptions, we and others (Dibattista & Gosse, 2006; Epstein & Brosvic, 2002) have found that students liked using the IF AT<sup>®</sup> forms and reported informally that it reflected what they learned better than standard multiple-choice formats. Therefore we can conclude that IF AT<sup>®</sup> forms are an effective and practical assessment tool and encourage more broad adoption by educators. More specifically, we encourage science/chemistry faculty who, in the author's experience, typically shy away from multiple-choice assessments because of the lack of ability to award partial-credit points. Finally, we would like to caution readers that, while our data is strongly indicative that our students' second-chance responses were grounded in learning, it is important that future research make efforts to probe the nature of learning triggered through immediate feedback in a testing situation.

### Acknowledgments

The authors thank Dr. Raymond Mugno in the Mathematics Department at Southern Connecticut State University. The authors wish to indicate no conflict of interest with Epstein Educational Enterprises, Inc.

### References

- Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: 7 research-based principles for smart teaching*. San Francisco, CA: Jossey-Bass. Retrieved from <http://c4ed.lib.kmutt.ac.th/sites/default/files/HowLearningWorks-Ambrose.pdf>
- Attali, Y. (2011). Immediate Feedback and Opportunity to Revise Answers Application of a Graded Response IRT Model. *Applied Psychological Measurement*, 35(6), 472-479. <http://dx.doi.org/10.1177/0146621610381755>
- Brosvic, G. M., Epstein, M. L., Cook, M. J., & Dihoff, R. E. (2005). Efficacy of Error for the Correction of Initially Incorrect Assumptions and of Feedback for the Affirmation of Correct Responding: Learning In the Classroom. *The Psychological Record*, 55(3), 401-418. Retrieved from <http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1193&context=tp>
- Bush, M. (2015). Reducing the need for guesswork in multiple-choice tests. *Assessment & Evaluation in Higher*

- Education*, 40(2), 218-231. <http://dx.doi.org/10.1080/02602938.2014.902192>
- Dibattista, D., & Gosse, L. (2006). Test Anxiety and the Immediate Feedback Assessment Technique. *The Journal of Experimental Education*, 74(4), 311-328. <http://dx.doi.org/10.3200/JEXE.74.4.311-328>
- Dihoff, R. E., Brosvic, G. M., & Epstein, M. L. (2003). The Role of Feedback During Academic Testing: The Delay Retention Effect Revisited. *The Psychological Record*, 53(4), 535-548. Retrieved from <http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1496&context=tpr>
- 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), 207-232. Retrieved from <http://www.if-at.com/home/articles/file/research/ImmediateFeedback.pdf>
- Epstein, M. L., Epstein, B. B., & Brosvic, G. M. (2001). Immediate Feedback During Academic Testing. *Psychological Reports*, 88(3), 889-894. <http://dx.doi.org/10.2466/pr0.2001.88.3.889>
- Epstein, M. L., & Brosvic, G. M. (2002). Students Prefer the Immediate Feedback Assessment Technique. *Psychological Reports*, 90(3c), 1136-1138. <http://dx.doi.org/10.2466/pr0.2002.90.3c.1136>
- Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2002). Immediate Feedback Assessment Technique Promotes Learning and Corrects Inaccurate First Responses. *The Psychological Record*, 52, 187-201. Retrieved from <http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1212&context=tpr>
- Fischer, G. (1999). Lifelong Learning: Changing Mindsets. In G. Cumming, T. Okamoto, & L. Gomez (Eds.), *7<sup>th</sup> International Conference on Computers in Education on "New Human Abilities for the Networked Society"* (pp. 9-10). Omaha, NE: IOS Press. Retrieved from <http://l3d.cs.colorado.edu/~gerhard/papers/1999-ICCE-final-submitted.pdf>
- Heaslip, G., Donovan, P., & Cullen, J. G. (2014). Student Response Systems and Learner Engagement in Large Classes. *Active Learning in Higher Education*, 15(1), 11-24. <http://dx.doi.org/10.1177/1469787413514648>
- Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-Choice Tests Exonerated, at Least of Some Charges Fostering Test-Induced Learning and Avoiding Test-Induced Forgetting. *Psychological Science*, 23(11), 1337-1344. <http://dx.doi.org/10.1177/0956797612443370>
- 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(2), 199-218. <http://dx.doi.org/10.1080/03075070600572090>
- Roediger, H. L. III., & Butler, A. C. (2011). The Critical Role of Retrieval Practice In Long-Term Retention. *Trends in Cognitive Sciences*, 15(1), 20-27. <http://dx.doi.org/10.1016/j.tics.2010.09.003>
- Scharf, E. M., & Baldwin, L. P. (2007). Assessing Multiple Choice Question (MCQ) Tests-A Mathematical Perspective. *Active Learning in Higher Education*, 8(1), 31-47. <http://dx.doi.org/10.1177/1469787407074009>
- Schell, J., Lukoff, B., & Mazur, E. (2013). Catalyzing Learner Engagement Using Cutting-Edge Classroom Response Systems In Higher Education. *Cutting-edge Technologies in Higher Education*, 6, 233-261. [http://dx.doi.org/10.1108/S2044-9968\(2013\)000006E011](http://dx.doi.org/10.1108/S2044-9968(2013)000006E011)
- Slepkov, A. D. (2013). Integrated Testlets And The Immediate Feedback Assessment Technique. *American Journal of Physics*, 81(10), 782-791. <http://dx.doi.org/10.1119/1.4820241>
- Smith, J. A. (2013). Immediate Feedback and Assessment Technique (IF AT<sup>®</sup>) Testing Forms: An Overview of the Tool and Uses. *Developments in Business Simulation and Experiential Learning*, 40, 107-109. Retrieved from <http://journals.tdl.org/absel/index.php/absel/article/view/28>

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).