

1-1-2012

# Empirical Evidence that Formative Assessments Improve Final Exams

Carol Springer Sargent  
*Macon State College School of Business*

Andrea A. Curcio  
*Georgia State University College of Law, [acurcio@gsu.edu](mailto:acurcio@gsu.edu)*

Follow this and additional works at: [https://readingroom.law.gsu.edu/faculty\\_pub](https://readingroom.law.gsu.edu/faculty_pub)

 Part of the [Legal Education Commons](#)

---

## Recommended Citation

Carol Springer Sargent & Andrea A. Curcio, Empirical Evidence that Formative Assessments Improve Final Exams, 61 J. Legal Educ. 379 (2012).

This Article is brought to you for free and open access by the Faculty Publications at Reading Room. It has been accepted for inclusion in Faculty Publications By Year by an authorized administrator of Reading Room. For more information, please contact [mbutler@gsu.edu](mailto:mbutler@gsu.edu).

# Empirical Evidence that Formative Assessments Improve Final Exams

Carol Springer Sargent and Andrea A. Curcio

## Introduction

Law school may be one of the few spots on campus still using a comprehensive exam for the entire course grade, even though many have called for an end to this single assessment model.<sup>1</sup> Pedagogical scholars have suggested that law professors begin using formative assessments, assessments designed to provide students with feedback throughout the semester, arguing that giving regular feedback enhances student learning and performance.<sup>2</sup> Students also believe they could learn better if they had more feedback, and many voice deep frustration at the low quality and quantity of feedback during the semester from their professors.<sup>3</sup> Despite the call for more feedback

**Carol Springer Sargent** is an Associate Professor, School of Business at Macon State College.

**Andrea A. Curcio** is a Professor of Law and Co-Director of the Externship Program, Georgia State University College of Law.

The authors thank Professors Leah Christensen, Harry Dangel, Eileen Kaufman and Leslie Wolf for their thoughtful comments on earlier drafts. For insights on study design and potential explanation of results, the authors are indebted to the Center for Teaching and Learning Fellows at Georgia State University and participants at the GTA Pedagogy Conference at Georgia State University. They also thank Gregory Todd Jones for his help designing and implementing the electronic student survey used in this study. Finally, they thank Cindy Perry for her tireless data entry efforts.

1. See Gregory S. Munro, Outcomes Assessment for Law Schools 143 (Inst. for L. Sch. Teaching 2000); Michael Hunter Schwartz, Sophie Sparrow & Gerald Hess, Teaching Law by Design: Engaging Students from the Syllabus to the Final Exam 154-58 (Carolina Acad. Press 2009); William M. Sullivan, Anne Colby, Judith Welch Wegner, Lloyd Bond & Lee S. Schulman, Educating Lawyers: Preparation for the Profession of Law 164-67 (Jossey-Bass 2007) [hereinafter *Educating Lawyers*].
2. Schwartz, *supra* note 1, at 135-37; Roy Stuckey and Others, Best Practices for Legal Education: A Vision and A Road Map 190-191 (Clinical Legal Educ. Ass'n 2007); *Educating Lawyers*, *supra* note 1, at 164-67.
3. *Educating Lawyers*, *supra* note 1, at 165; see also Janice Orrell, Feedback on Learning Achievement: Rhetoric and Relating, 11 Teaching in Higher Educ. 441, 441 (2006) (citing to studies in other disciplines).

and guiding assessments in legal education, the single end-of-semester exam remains the norm,<sup>4</sup> so the impact of formative assessments on law students' learning remains largely unexplored.

Of course, there are good reasons to prefer a single cumulative final exam. For example, grading multiple assessments may not be realistic given professors' other commitments.<sup>5</sup> The end-of-semester comprehensive exam may also be the best measure of whether students see the "big picture" and have successfully integrated the various doctrinal materials.<sup>6</sup> Periodic testing on the individual parts may not represent the whole-cloth thinking needed for law student development and thus mislead them into studying the parts rather than how they interrelate.

Other disciplines have struggled with how to encourage wide-scale adoption of best assessment practices.<sup>7</sup> The limited use of formative assessments may be partly attributable to the fact that some law professors, like educators in other disciplines, are unaware of the value of formative assessments because study results are published in educational journals generally not read by law professors. Even those who do offer formative assessments sometimes get frustrated when students fail to use or don't understand the feedback provided,<sup>8</sup> so not all formative assessment practices implement perfectly. Policy makers' reluctance to tell teachers what to do in their classroom,<sup>9</sup> and the culture of academic freedom, makes imposing formative assessments unlikely.<sup>10</sup> Thus, the use of formative assessments depends upon professors' views about the value of this type of assessment.

This study provides evidence that formative assessments help law student performance on a cumulative final exam, and addresses some of the concerns expressed about integrating formative assessments into large-section doctrinal courses. The study builds upon prior work in a first-year civil procedure

4. Educating Lawyers, *supra* note 1, at 162.

5. Ruth Colker, Extra Time as an Accommodation, 69 U. Pitt. L. Rev. 413, 464 (2008).

6. See John D. Schunk, Can Legal Writing Programs Benefit from Evaluating Student Writing Using Single-Submission, Semester-Ending, Standardized, Performance Type Assignments?, 29 Hamline L. Rev. 308, 323 (2006).

7. See Paul Black & Dylan Wiliam, In Praise of Educational Research: Formative Assessment, 29 Brit. Educ. Res. J. 623 (2003) [hereinafter In Praise of Educational Research]; Orrell, *supra* note 3, at 441; Alison Rushton, Formative Assessment: A Key to Deep Learning?, 27 Med. Tchr. 509 (2005).

8. Orrell, *supra* note 3, at 441 (citing studies in other disciplines in which professors expressed frustration at students' failure to utilize the professor's feedback); Aida M. Alaka, Phenomenology of Error in Legal Writing, 28 Quinnipiac L. Rev. 1, 1 at n.13 (2009) (noting legal writing professors' frustration when students fail to incorporate professor feedback into subsequent assignments).

9. In Praise of Educational Research, *supra* note 7, at 629-30 (discussing the issue of imposing formative assessment practices in context of teaching secondary school).

10. See *id.* at 633-34.

course which found that practice essay questions improved the performance of students with above-the-median LSAT scores and above-the-median UGPAs on a traditional cumulative final exam.<sup>11</sup> Instead of using longer take-home essay assignments, the current work used both in and outside of class short-essay and short-answer ungraded quizzes, a graded mid-term, and self-reflective exercises to examine whether a different set of formative assessments impacts students' cumulative final exam performance, and, if so, whether that impact was across-the-board or confined to a particular segment of students.

### Formative Assessments

Formative assessments seek to increase learning and motivation by offering students feedback about gaps between current and desired levels of performance.<sup>12</sup> Summative assessments, by contrast, seek to measure the amount of learning.<sup>13</sup> These descriptors, formative and summative, indicate the purpose rather than the content of the materials.<sup>14</sup>

The effectiveness of feedback from formative assessments depends on (1) what you give students and (2) the way students receive or interpret it. Examining first what you give students, the literature shows that feedback comes in a variety of formats and sizes, and that even in small quantities, most feedback improves student performance.<sup>15</sup> Feedback allows learners to calibrate their progress towards academic goals.<sup>16</sup> The effect is greater when the feedback offers an explanation rather than just a correct response,<sup>17</sup> and

11. Andrea A. Curcio, Gregory Todd Jones & Tanya M. Washington, Developing an Empirical Model to Test Whether Required Writing Exercises or Other Changes in Large-Section Class Teaching Methodologies Result in Improved Exam Performance, 57 J. Legal Educ. 195, 197 (2007) [hereinafter Developing an Empirical Model] (briefly explaining the study and its results); Andrea A. Curcio, Gregory Todd Jones & Tanya M. Washington, Does Practice Make Perfect? An Empirical Examination of the Impact of Practice Essays on Essay Exam Performance, 35 Fla. St. U. L. Rev. 271, 280-82, 302-306 (2008) [hereinafter Does Practice Make Perfect] (providing an in-depth explanation of metacognitive theory and the role the authors suggest metacognition may have played in the study's results).
12. Schwartz, *supra* note 1, at 137; Rushton, *supra* note 7, at 509; D. Royce Sadler, Formative Assessment and the Design of Instructional Systems, 18 Instructional Sci. 119, 120-21 (1989).
13. Stuckey, *supra* note 2, at 176.
14. In Praise of Educational Research, *supra* note 7, at 623.
15. Robert L. Bangert-Drowns, Chen-Lin C. Kulik, James A. Kulik & Mary Teresa Morgan, The Instructional Effect of Feedback in Test-Like Events, 61 Rev. Educ. Res. 213, 215 (1991); In Praise of Educational Research, *supra* note 7, at 629; *see generally* Valerie J. Shute, Focus on Formative Feedback, 78 Rev. Educ. Res. 153 (2008) (reviewing numerous studies on the effect of formative feedback and discussing which feedback methods produce positive learning effects).
16. Stuckey, *supra* note 2, at 93.
17. Bangert-Drowns, *supra* note 15, at 232.

is available immediately after performing,<sup>18</sup> although there are also some benefits to delayed feedback.<sup>19</sup> Feedback is more effective when it provides details of how to improve<sup>20</sup> and explains why an answer is correct.<sup>21</sup> Feedback interventions for complex tasks yield weaker effects than those for simpler tasks because tasks requiring little cognitive effort are easy to remedy with motivation alone.<sup>22</sup> Numerous studies suggest that feedback may be more effective if ungraded because students tend to focus on grades, not suggestions for improvement.<sup>23</sup> Feedback that directs students' attention to themselves, either because it contains praise or negative comments, may be less effective in improving student performance or learning because it directs attention away from the task at hand.<sup>24</sup> Overall, while content and timing play a part, most feedback helps learning.

Turning to student responses to the feedback given, formative assessments do not just fix misconceptions and knowledge lapses, they potentially change student motivation and study strategies. By giving students information about shortfalls early in the course, they have the opportunity to adjust and improve, potentially inspiring more effort.<sup>25</sup> Feedback can also suppress motivation. For example, formative assessments that compare a student's performance to that of his or her classmates may inhibit learning because when faced with such comparisons "people who perform poorly tend to attribute their failures to lack of ability, expect to perform poorly in the future, and demonstrate decreased motivation on subsequent tasks."<sup>26</sup> Not surprisingly, feedback promotes

18. James A. Kulik & Chen-Lin C. Kulik, Timing of Feedback and Verbal Learning, 58 Rev. Educ. Res. 79, 89 (1988).
19. Shute, *supra* note 15, at 163-66 (discussing and summarizing various studies about the positive and negative effects of both immediate and delayed feedback).
20. *Id.* at 157-58 (noting that studies indicate that specific feedback is more effective and also cautioning that specificity alone does not necessarily explain a particular feedback intervention's effect).
21. Bangert-Drowns, *supra* note 15, at 232; Avraham N. Kluger & Angelo De Nisi, The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory, 119 Psychol. Bull. 254, 265, 268 (1996).
22. Kluger, *supra* note 21, at 268-69.
23. Paul Black & Dylan Wiliam, Assessment and Classroom Learning, 5 Assessment in Educ.: Principles, Pol'y & Prac. 7, 23 (1998) (reviewing the literature and noting that use of overall grades in conjunction with feedback can have a detrimental effect).
24. Kluger, *supra* note 21, at 267; Shute, *supra* note 15, at 169 (reporting results from numerous studies finding that feedback containing praise may not be an effective way to enhance learning and performance).
25. Paul R. Pintrich, A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts, 95 J. Educ. Psychol. 667, 672 (2003).
26. Shute, *supra* note 15, at 167 (explaining research findings by Kluger and DeNisi, discussed in greater detail *supra* note 21).

learning best if it is received mindfully,<sup>27</sup> if students accurately perceive what they do not know,<sup>28</sup> and if they are motivated to fix the problem.<sup>29</sup> In summary, not all students are helped equally<sup>30</sup> because feedback effectiveness turns not just on the materials provided, but also on the ability of the recipient to digest and use the feedback,<sup>31</sup> as well as their goals, self-confidence, interest, and intentions.<sup>32</sup>

### Formative Feedback for Law Students

While many have called for the introduction of formative assessments into law school classes,<sup>33</sup> few studies have sought to determine whether this kind of feedback has verifiable advantages for students.<sup>34</sup> One earlier study in a first-year civil procedure class found that multiple practice essay questions followed by annotated model answers helped students with above-the-median LSAT scores and UGPAs break down legal rules and perform a complex factual analysis on a final exam.<sup>35</sup> The current work builds on that study which compared two sections of a required civil procedure course, one

27. Gavriel Salomon & Tamar Globerson, Skill May Not Be Enough: The Role of Mindfulness in Learning and Transfer, 11 *Int'l J. Educ. Res.* 623, 624 (1987).
28. John Biggs, Assessment and Classroom Learning: A Role for Summative Assessment? 5 *Assessment in Educ.: Principles, Pol'y and Prac.*, 103, 104 (1998).
29. *Id.* at 104; Shute, *supra* note 15, at 162.
30. *See* Does Practice Make Perfect, *supra* note 11 (finding that practice essay questions only helped students with above-the-median LSAT scores and above-the-median UGPAs).
31. *See infra* text accompanied by notes 78–80 (discussing students' varying metacognitive abilities).
32. *See infra* text accompanied by notes 76–77 and 81–88 (discussing the interaction between feedback and other factors that may make feedback more or less effective).
33. *See e.g.*, Educating Lawyers, *supra* note 1, at 171; Munro, *supra* note 1, at 151; Schwartz, *supra* note 1, at 136, 155–58.
34. There have been studies on the impact of teaching methodologies on student learning: *see e.g.*, Robin A. Boyle, Karen Russo & Rose Frances Lefkowitz, Presenting a New Instructional Tool for Teaching Law-Related Courses: A Contract Activity Package for Motivated and Independent Learners, 38 *Gonz. L. Rev.* 1 (2003); Paul F. Teich, Research on American Law Teaching: Is There A Case Against the Case System?, 36 *J. Legal Educ.* 167 (1986) (reviewing empirical studies on the impact of different law school teaching methodologies on student exam performance and reviewing teaching methodology studies from other disciplines); William K. S. Wang, The Restructuring of Legal Education Along Functional Lines, 17 *J. Contemp. Legal Issues* 331, 353 n.49 (2008) (reviewing existing empirical studies on the impact of technology on law student learning). Additionally, at least one professor has noted that subjectively, he believed his law students performed better after engaging in practice writing exercises. John M. Burman, Out-of-Class Assignments as a Method of Teaching and Evaluating Law Students, 42 *J. Legal Educ.* 447, 453 (1992). However, other than the study conducted by Curcio et al., *supra* note 11, the authors could find no other empirical study specifically addressing the impact of formative assessments on law student performance.
35. Does Practice Make Perfect, *supra* note 11.

giving students practice essay exam questions accompanied by various forms of feedback throughout the semester, and a control section with no formative assessments.<sup>36</sup> One weakness of the study was that, although both classes took the same traditional cumulative final exam, they were taught by different instructors.<sup>37</sup> The present investigation eliminates the instructor confound because the same instructor taught both groups.

One surprise from the prior work was that practice essays only helped students with above-the-median LSAT scores and above-the-median UGPAs.<sup>38</sup> The study's authors attributed this to students' different metacognitive abilities,<sup>39</sup> i.e. students' ability to identify what they did wrong (self-observe and self-judge) and develop and implement strategies to fix their weaknesses (self-react).<sup>40</sup> The cognitive psychology literature claims that students with stronger metacognitive skills are better able to use the information gleaned from feedback and apply it to new exam question scenarios and that these abilities can be developed.<sup>41</sup> The present study adds self-reflective exercises to help students understand the specific gaps in their analysis and reasoning, hopefully strengthening their ability to self-observe, self-judge, and self-react.

The current work strengthens the statistical methods used to detect learning differences. The prior work used a t-test to examine differences in exam scores between the two groups, one with practice essays (intervention) and the other without (control).<sup>42</sup> Although the t-test identified a statistical difference between the two groups, it could not identify how much of the difference was associated with any particular causal variable—in other words, how much of the exam score difference was attributable to the intervention versus other variables such as law school predictors or other academic behaviors. In this study, we use a regression model allowing us to look at variables other than the formative assessments that might predict the exam score differences, such as motivation to use feedback, law school grades, or law school grade predictors.

36. *Id.* at 286-87.

37. *Id.* at 286.

38. *Id.* at 293-97.

39. *Id.* at 302-306.

40. Dale H. Schunk, *Metacognition, Self-Regulation, and Self-Regulated Learning: Research Recommendations*, 20 *Educ. Psychol. Rev.* 463, 465 (2008) (explaining meta-cognition).

41. Sofie M. M. Loyens, Joshua Magda & Remy M. J. P. Rikers, *Self-Directed Learning in Problem-Based Learning and its Relationships with Self-Regulated Learning*, 20 *Educ. Psychol. Rev.* 411 (2008) (discussing studies investigating the role of self-directed and self-regulated learning in problem-based learning scenarios); *see also* Michael Hunter Schwartz, *Expert Learning for Law Students* (Carolina Acad. Press 2005) (discussing how law students can develop self-regulated learning skills and thus improve their learning and performance).

42. The t-tests compared the mean raw exam score between the two groups. A t-test does not reveal which variable or variables caused the difference. Because students in both the control and intervention group had equivalent LSAT scores and UGPAs, the prior study claimed the formative assessments accounted for the exam score difference.

For example, one might predict that any difference in final exam scores between the control and intervention groups would be due to students' LGPAs and not just the intervention. The regression analysis permits variation in exam scores expected from sources other than the intervention, such as varying academic ability (grades), LSAT and completion of reflective activities, to be adjusted in the model before comparing the final exam scores of the two groups. In other words, the regression model allows us to control for exam score predictors and view the benefit of the interventions separate from other exam predictors.

### Hypotheses

Because feedback improves students' ability to self-observe, self-judge and self-remedy; fuels motivation; and corrects misconceptions early in the learning process; we hypothesize: Students receiving formative assessments will have higher final exam scores than students in the control group.

Because the effectiveness of feedback is dependent on student motivation and cognitive ability to use the feedback, we hypothesize: Students with above-the-median LSAT scores and above-the-median UGPAs and LGPAs will have a greater advantage from formative assessments.

### Method

#### *Participants*

Participants were students enrolled in a second-year required Evidence course in spring 2008 (n=67) and spring 2009 (n=51),<sup>43</sup> both taught by the second author at a second-tier urban public law school with a diverse student body. The professor has taught this course, with minor adjustments for changes in law, since 1996.

#### *Procedure*

The 2008 Evidence students were the control section. They were taught using a problem method supplemented by case analysis. Those students had one cumulative final exam counting as the full course grade. The 2009 Evidence students, the intervention group, were taught using the problem method supplemented by case analysis, but also received a series of formative assessments, including five ungraded quizzes and a graded midterm.<sup>44</sup>

43. There were fifty-two students enrolled in the 2009 class. One student did not turn in a consent form and thus was not included in the study data.
44. Because the literature suggests that feedback that is graded, norm-referenced, and contains comments that direct attention to self may be unhelpful (*see supra* text accompanying notes 23-26), the vast majority of feedback in this course was given in the form of model answers, grading rubrics and self-reflective analysis. The mid-term was graded so that the students had the opportunity to compare their self-analysis to the professor's. At their request, students also were given information about the class median in order to have a sense of how



Following the quizzes and mid-term, students were given model answers and grading rubrics and were asked to engage in reflective exercises to help them calibrate their comprehension and prepare for the cumulative final exam. The final exam in the intervention section counted 83 percent of their course grade because the professor wanted the course grade to correspond primarily to the summative assessment, the comprehensive final.

All students, control and intervention, were taught using the same casebook and other materials. The only difference in substantive coverage was that students in the intervention group had about five hours less in-class case and hypothetical problem analysis so that they could complete or review quizzes, reflective exercises, and the midterm. Thus, the depth of in-class coverage was somewhat more in the control group, although the breadth of coverage remained the same.

The second author wrote the practice test questions, model answers, grading rubrics, and self-reflective exercises.<sup>45</sup> The self-reflective exercises were an attempt to ensure that the feedback was received mindfully<sup>46</sup> and the exercises became more extensive as the semester progressed. Initially, students were asked general questions such as “did you identify the correct legal issue?” However, the professor’s review of the self-reflective exercises indicated that some students did not seem to understand that they had made errors.<sup>47</sup> Even with a model answer available to them, some students said that they had identified the correct issue when in fact they had not. Thus, as the semester progressed, the self-reflective questions became more directed.<sup>48</sup>

In the intervention group, at the end of weeks one and three, students were given an in-class ungraded timed quiz, followed by class time to compare their responses to a model answer, a grading rubric,<sup>49</sup> and a brief self-reflective

---

their mid-term grade compared to those of their colleagues.

45. For examples of quiz and self-reflection questions, see Appendices A and B.
46. See Salomon, *supra* note 27, at 624 (discussing the role of mindfulness in learning feedback).
47. Accurate self-reflection skills are not necessarily intuitive and students need to be taught how to develop these skills. See generally Sadler, *supra* note 12 (arguing student self-reflective skills can be developed by providing authentic evaluative experiences); see also Malathi Srinivasan et al., Does Feedback Matter? Practice-Based Learning for Medical Students After a Multi-Institutional Clinical Performance Examination, 41 *Med. Educ.* 857, 858 (2007) (reporting results from studies indicating that “[m]edical student self-assessment of clinical and cognitive skills . . . correlate weakly ( $r=0.00-0.25$ ) or moderately ( $0.25-0.50$ ) with expert ratings of student performance.”).
48. For example, by the end of the semester, the self-reflective materials included directions such as: “This question involved Rule 702, underline where in your answer you identified Rule 702 as the controlling rule.” For other examples of self-reflective questions, see *infra* notes 51-52, 55; see also Appendix B.
49. With the exception of the first ungraded quiz, all other formative assessments included grading rubrics so that the students could see how the professor allocated points indicating the relative importance of each component of the answer.

questionnaire.<sup>50</sup> The professor asked the students to turn in their self-reflective questionnaire but they were not required to do so.<sup>51</sup> At the beginning of week five, students took a three-question ungraded short-answer take-home quiz, which was due the next class. Again, students had class time to review a model answer and grading rubric and were asked to complete a longer self-reflection. Students received an extra raw score point toward their total grade if they turned in this self-reflection.<sup>52</sup> At the end of week seven, the professor gave students a multi-issue in-class un-graded quiz accompanied by a model answer, grading rubric, and self-reflective questionnaire. For this quiz, students were asked to peer edit a classmate's responses during class and self-edit outside of class. At the end of week eight, students took an open-book in-class timed graded mid-term consisting of one multi-issue short essay question (350 words) and three short answer questions (150 words).<sup>53</sup> At the end of week ten, the professor returned the students' mid-term with comments,<sup>54</sup> a grading rubric, and a model answer which also contained information about common errors and how to avoid those in the future. Students were asked to grade themselves using the rubric and to complete a set of self-reflective questions before they were given their graded exam answers.<sup>55</sup> At the beginning of the last week of class, week fourteen, students received the last take-home ungraded quiz with one multiple-issue short essay question and two short answer single issue

50. For an excerpt of questions from the first quiz, *see* Appendix A.
51. Thirty-four out of fifty-two students turned in the self-reflective analysis questions. Many responses were "yes" and "no" answers to questions such as: "Did you spot the proper issue? How did you know it was the issue? If you didn't spot the issue, what can you do to help yourself do this next time?" It is difficult to know if the students absorbed the self-monitoring that the question was intended to prompt but didn't take the time to write out a full response, since the activity was clearly for the students' benefit only.
52. Forty-three of fifty-two students turned in the self-reflective analysis questions. Because the students got one raw score point for a good faith effort, all who answered these questions provided some explanation to questions such as: 1. "This question required you to use 404(b) (prior bad acts) as the basis of your analysis. Did you identify that rule [at least to yourself]? What words in this problem clue you in to the fact that you have to look at Rule 404(b)?...3. Did you explain how this evidence would go to the defendant's knowledge that Dylan would take the drugs (and explain HOW it went to that knowledge)? Specifically what facts did you use to support your argument that it went to her knowledge that Dylan would take the drugs? Were there any others you could have used?"
53. For examples of the short answer mid-term questions, *see* Appendix A.
54. Generally, the comments were directed toward areas where the students needed to improve: e.g. "you need to practice issue spotting."
55. Students were asked questions such as, "Did you explain how the probative value is low? If so, identify where in your answer you explained how the probative value was low and identify the facts you used to support your argument." It also asked students to compare the grade given by the professor with the grade they gave themselves. Most students noted that the grades they gave themselves using the rubric were close to the grade given by the professor. A sample of some of the self-reflective questions accompanying the graded mid-term are found in Appendix B.

questions. A model answer and grading rubric were also distributed along with a very short set of self-reflective questions.

At the end of the intervention section (but before the final exam was taken), students completed a survey reporting their opinion about the impact of the various formative assessments on their learning.<sup>56</sup>

## Measurements

### *Prior Achievement*

Students' undergraduate grade point average (UGPA) was used to measure prior general academic achievement. Law school grade point average (LGPA) was used to measure achievement in the first year of law school. These variables were used to predict exam performance in the regression model.<sup>57</sup> We also used these measures to replicate the prior study where those with weaker law school predictors failed to benefit from the formative assessments.<sup>58</sup>

### *LSAT Scores*

Our analysis indicates that for students in the control and intervention groups, LSAT score was not correlated with undergraduate grades (Pearson correlation = 0.015;  $p = 0.871$ ),<sup>59</sup> first year law school grades<sup>60</sup> (Pearson correlation = 0.148;  $p = 0.122$ ) or performance in the second year course in this study (Pearson correlation = 0.087;  $p = 0.350$ ). However, the LSAT scores were a predictor of responsiveness to formative assessments in the prior work so we used them to replicate the earlier research.

56. For a discussion of students' views of the impact of the exercises on their learning, *see infra* text accompanying note 72; for a visual representation of students' views, *see infra* Figures 2 and 3.
57. For a discussion of how the regression model worked, *see infra* text and accompanying notes 63-67.
58. Does Practice Make Perfect, *supra* note 11.
59. This finding was consistent with an earlier study at the same school involving students who would have taken Evidence in 2007. *Id.* at 293.
60. The lack of correlation between first-year grades and LSAT scores in both the 2008 and 2009 cohorts involved in this study is noteworthy both because it is contrary to findings in another study at this same school done with students who would have taken Evidence in 2007—one year before the 2008 cohort involved in this study, *id.* (finding a Pearson Correlation between LSAT scores and overall spring LGPA of .286 with a Sig (2-tailed) .001) and because it is contrary to findings in other studies, *see e.g.*, Abiel Wong, Note, "Boalt-ing" Opportunity?: Deconstructing Elite Norms in Law School Admissions, 6 *Geo. J. of Poverty L & Pol'y* 199, 227 (1999) (noting that "The LSAT's correlation coefficient with first year grades ranges from .01 to .62, depending on the law school, with a median correlation at .41"). The authors have no explanation as to why, for both the 2008 and 2009 law students involved in this study, there was no correlation between LSAT scores and first year law grades.

*Final Exam*

In both years, students were given a six-page “case file” prior to the exam and were told that a significant portion of the exam questions would come from evidentiary issues embedded in that case file.<sup>61</sup> Both exams were closed-book timed exams given during the designated exam period with equivalent time given to answer each question. Both exams were comprehensive final exams, testing all substantive areas covered throughout the semester.

Both exams had two short essay questions (of approximately 400 words each). The essay questions in 2008 were similar, but not identical to, those used in 2009. The remaining questions were eighteen (in 2008) or fifteen (in 2009) short answer questions (approximately 125–150 words).<sup>62</sup> Eleven short answer questions, worth four or five raw points each, were the same between the two years. The raw score on the common eleven short answer questions (common questions score) were used to test the hypotheses. The common question scores were highly correlated with the total final exam score (see Table 1).

The eleven common questions were graded using the same rubric for both years. To establish grading reliability, the administrative staff randomly selected fifteen exams from 2008 and fifteen exams from 2009 for blind grading (not knowing initial grade) by the second author. The eleven common questions were re-graded without knowing the original grade. The result of that blind grading exercise was that 295 out of 330, or 89.4 percent of the questions received the same point score as the original grade, 29 had a one raw score point difference (8.8 percent) with only 6 (1.8 percent) re-graded scores being off from the original score by more than one point.

*Extra Point for Self Reflective Exercises*

Students in the intervention section were given one point for turning in their self-reflection exercises in week five of the course. Nine students did not turn in these materials. The regression model included a variable indicating whether they submitted the self reflective exercises to control for motivation or interest in formative assessments.

61. For an explanation of how case files can be used to test Evidence students, see Andrea A. Curcio, Evidence Exams Using a Case File, AALS Evidence Sec. Newsl., Spr./Sum. 2009, at 8–9.
62. Students in 2008 had three more short answer questions than students in 2009. The 2008 students were given an additional half hour to answer the exam questions.

### Analysis

The data met the assumptions of normality<sup>63</sup> and the correlations between LSAT, UGPA and LGPA (Table 1) were low enough to include all three variables in the regression model.<sup>64</sup>

**Table 1**  
**Correlations Between Variables**

	LSAT	UGPA	LGPA	Final Exam Score
UGPA	0.015			
LGPA	0.148	0.192*		
Final Exam Score	0.087	0.147	0.503**	
Common questions score	0.010	0.145	0.371**	0.878**

\* Correlation is significant at the .05 level

\*\* Correlation is significant at the .01 level

Attributes of participants in the control and intervention sections, shown in Table 2, indicate no differences except for scores on the common exam question score.

**Table 2**  
**Participant Attributes: Mean (Std. Dev.)**

Attribute	Control	Intervention
Number of participants	67	51
Percent female	53.73%	50.98%
Percent Caucasian	71.64%	66.67%
Undergraduate GPA	3.26 (0.54)	3.32 (0.30)
LSAT	159.28 (3.15)	159.12 (2.75)
First Year Law School GPA*	78.87 (4.14)	79.00 (3.66)
Common Questions score**	31.64 (8.87)	34.73 (6.78)

\* We did not have law school grades for eight transfer students, four in the control section and four in the intervention section

\*\* Significant difference at < .05 level

63. The validity of the regression analysis depends on data having a bell shape, with most points clustering near the mean and fewer at the extreme ends, and the shape approaching symmetry about the mean. The data conformed to this requirement and hence met the assumptions of normality. There was one outlier, a LSAT score in the 2008 control section that was 3.4 standard deviations above the mean, which was retained as a valid data point. Removing the outlier did not change the results.
64. Regression does not work well when explanatory variables (i.e., LSAT, UGPA, LGPA) measure similar constructs and as such are overlapping measures. So, the correlation of variables must be checked prior to running a regression analysis to be sure they are low enough not to interfere with the model results. The correlation between UGPA, LGPA and LSAT scores was low enough to include all in the model.

### Formative Assessment Experience Impacting Performance

Formative assessment experience (the intervention) was significant in explaining the variance in common question scores in a regression with those scores as the outcome variable; formative assessments as the explanatory variable; LGPA, UGPA, and LSAT as controls for prior achievement; and “extra point” as a measure of motivation to use formative assessments ( $t=1.982$ ,  $p=0.050$ ).<sup>65</sup> The effect size,  $3.024$ <sup>66</sup> points out of 50 points, is about a half a letter grade (6.048 percent), a moderate effect size. This analysis omitted eight students (four in the control group and four in the intervention group) who transferred from another law school and therefore had no LGPA. Replacing the missing LGPA scores with the median LGPA so that the eight transfer students were not eliminated from the analysis, formative assessment experience was still significant in explaining common questions scores ( $t=2.115$ ,  $p=0.037$ ).

### Formative Assessment: Stratifying by LSAT Score and UGPA

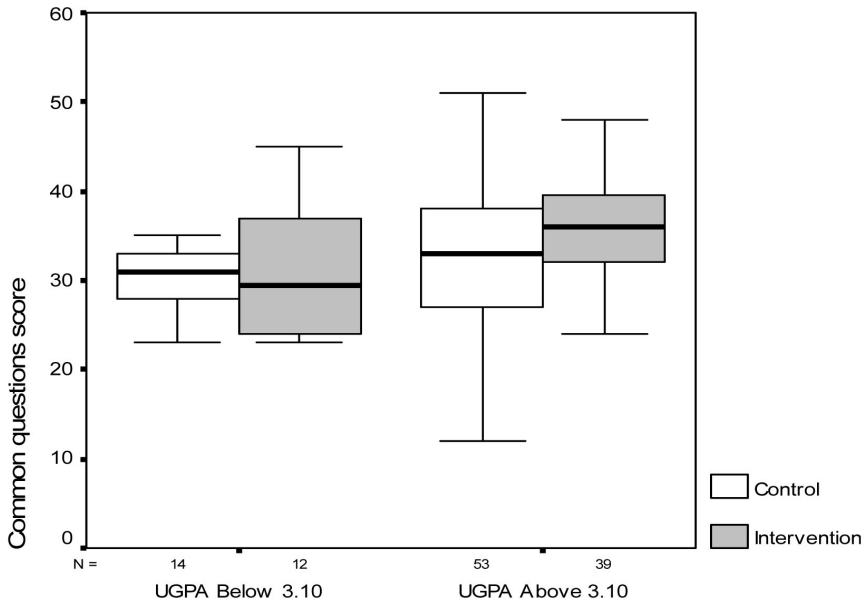
In order to test the second hypothesis, we split the participants into those below the median LSAT (159) and those at or above it. As hypothesized, we found that formative assessment experience (the intervention) was significant in explaining the variance in common question scores for students in the intervention group with at or above the median LSAT scores ( $t=2.539$ ,  $p=0.013$ ) but not for those below the median ( $t=-0.280$ ,  $p=0.782$ ).<sup>67</sup> The effect size for the top two-thirds of the intervention class,<sup>68</sup> 4.595 points out of 50 or almost a full letter grade (9.19 percent), is moderate to large. The results were similar with transfer students included in the model (by replacing the missing LGPA with median LGPA).

Re-running the above regressions with the participants split by roughly the same proportion of students, top two-thirds by LGPA and UGPA,<sup>69</sup> we found that experience with formative assessments was significant only in predicting common question scores for students in the top two-thirds of the class based

65. Using the regression analysis enabled us to determine it was likely the intervention, not students' LGPA, UGPA, LSAT score, or their motivation to use reflective activities that accounted for the difference in scores between the control and intervention students.
66. The regression equation indicates the predicted number of exam points attributed to each variable in the model (the *B*), permitting an effect size for each variable.
67. Because twenty-one students who had an LSAT score of 159 were put in the above-the-median group, only seventeen students in the intervention group fell below the median LSAT score.
68. Determined by the *B* in the regression equation for the intervention variable. See *supra* notes 66-67 for explanation of how the *B* in the regression equation works.
69. The groups were not split exactly at two-thirds and one-third. We split between clusters of nearly identical scores to avoid having students with nearly identical measures split between groups. This resulted in 67 percent of the intervention students in the top two-thirds of LSAT and 70 percent of the intervention students in the top two-thirds by UGPA.

on UGPA ( $t=2.202$ ,  $p=0.030$ ). For students in the top two-thirds of LGPA, the result approached significance ( $t=1.807$ ,  $p=0.074$ ). Figure 1 shows the distribution of scores between the two groups for the top two-thirds and the bottom one-third based on UGPA.

**Figure 1<sup>70</sup>**  
**Scores on Common Questions Shown by UGPA and Treatment Group**

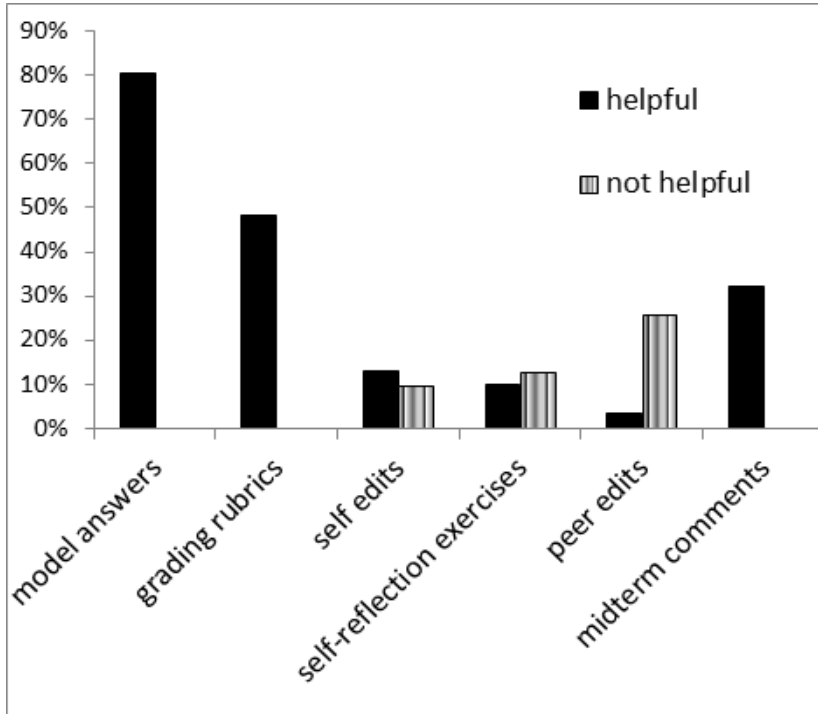


In the end-of-the-course survey ( $n=42$ ), intervention students, using a four point Likert scale, reported how helpful the formative assessments and accompanying feedback were to their learning (Figure 2). While many students found all the formative materials helpful, the graded midterm was rated as helpful or very helpful more than the other exercises. Over a third of the students who responded to the survey found the self-reflective exercises unhelpful. Thirty-one students in the intervention group added comments in the open-ended section of the end-of-the-course survey (Figure 3) indicating that model answers, grading rubrics, and professor comments were the most helpful and the peer edits and self reflections were the least useful feedback features.<sup>71</sup>

70. Box plots are a way to illustrate the range of scores. The range from first quartile to third quartile forms the “box” and the line in the middle of the box is the median score. The brackets above and below the box show the full range of scores.

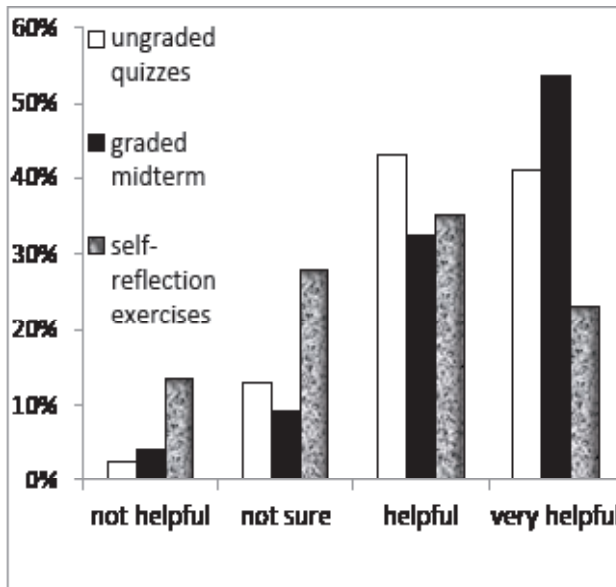
71. While students’ own reports of helpfulness gives insight about their perception, there was no direct measure of whether the individual components (self-reflective exercises, quizzes or midterms) were in fact helpful or not.

**Figure 2**  
**Ratings of Feedback and Self-Reflection Activities**





**Figure 3**  
**Survey Comments on Formative Assessment Activities**



### Discussion

Comparing the control and intervention scores on the common final exam questions, formative assessments improved performance for a majority of students taking a second year Evidence course. Splitting the class into the top two-thirds and bottom one-third by LSAT scores and UGPAs, we found that the effect was concentrated with students in the top two-thirds, *regardless of their first-year law school grades*.<sup>72</sup> Thus, the benefit inured to students with both above and below the median law school first year grades. The effect size for those who benefitted was moderate to large, just over nine percentage points.

The prior study found that practice essays helped those with above median law school predictors (UGPA/LSAT).<sup>73</sup> This work replicates the prior study and expands the effect to a large portion of the class (two-thirds). The current work also goes beyond the prior study, which only looked at grade predictors, not actual law school achievement. Thus, it did not look at whether the formative assessments helped students with below-the-median law school grades. This study demonstrated that formative assessments had a positive impact on students with below-the-median first year law school grades, as long as those students were not in the bottom one-third of the class in terms of either UGPA or LSAT score.

72. In our sample, LGPA was correlated with UGPA, but there was no correlation with LSAT scores (see Table 1). Thus, students may have had below the median law school grades and above the median LSAT scores or above the median UGPAs.

73. Does Practice Make Perfect, *supra* note 11.

In the hope of expanding the reach of the formative assessments to a larger segment of the class, this work varied from the earlier study by using a series of short essay and short answer practice exam questions instead of longer essay questions, providing time in class to complete the assessments, including a grading rubric with all but one of the formative assessments, and adding self-reflective question exercises after all assessments. However, we do not know specifically which of the new resources prompted higher exam scores in the intervention section. For example, we do not know whether the self-reflective exercises were a significant contributor to the improvement in students' performance. Students certainly seemed to believe that the self-reflective exercises were not particularly helpful.<sup>74</sup> Of course, their perceptions are not a direct measure of the actual helpfulness of the materials.

What is discernible from the study's results is that 70 percent of the intervention group benefitted substantially (nearly a letter grade) from the formative assessment materials. Unfortunately, 30 percent did not or could not use the materials to monitor and improve the quality of their work against instructor standards. There are a number of potential explanations for why some students benefited more than others from the practice tests and self-reflective exercises.

One reason stems from the fact that not all students are able to use feedback to improve. Information on gaps between current performance and desired standards is considered feedback "only if used to alter the gap."<sup>75</sup> LSAT scores and UGPA may reflect experience with successfully using feedback to improve test scores and an ability to use feedback to narrow achievement shortfalls. As such, those with higher LSAT scores and UGPAs may be more experienced with adjusting their study habits, re-working content units, organizing their answers to mirror the models given, and reflecting on gaps between attained and desired levels of performance.<sup>76</sup> Accordingly, the disproportional boost they enjoyed from the formative assessments reflects their greater ability to use feedback messages to correct and enhance understanding.

Another reason for the lack of effect for the lower scoring LSAT/UGPA students may lie in difficulty in perceiving the feedback messages or calibrating their comprehension. Students' ability to identify what they know

74. See *supra*, Figure 2 and Figure 3.

75. Sadler, *supra* note 12, at 121.

76. Barry J. Zimmerman, Self-Regulated Learning and Academic Achievement: An Overview, 25 *Educ. Psychologist* 3, 4-5 (1990) (noting that students with strong metacognitive skills can "plan, set goals, organize, self-monitor and self-evaluate at various points during the process of acquisition" and that doing so allows them to be "self-aware, knowledgeable, and decisive in their approach to learning"); see also Sadler, *supra* note 12, at 121 (describing the process necessary to effectively use feedback); Does Practice Make Perfect, *supra* note 11, at 280-282; 302-305 (discussing why law students with higher LSAT scores and UGPAs may have stronger meta-cognitive skills and hence be more able to use feedback to improve exam performance).

and don't know is a meta-cognitive skill.<sup>77</sup> LSAT scores and UGPA may reflect stronger meta-cognitive abilities and thus stronger abilities to identify lapses in knowledge and understanding.<sup>78</sup> In a study of medical students, most were initially poor at self assessment but improved with feedback.<sup>79</sup> This raises the question of whether these self-evaluative skills can improve with sufficient practice for law students as well. Perhaps students with lower LSAT/UGPAs, at least those interested in formative assessments, might get better at calibrating their comprehension and therefore turn into stronger law students if enough of their courses offered formative materials. However, in this course, the materials may have been too little or the wrong kind to help those underdeveloped in self-monitoring practices. Additionally, the self-reflection exercises were very brief at the start of the semester and became more extensive and directed as the course progressed. A longer and more developed set of self-reflection exercises earlier in the semester may have increased the effect size or helped those with lower LSAT scores/UGPAs scores more.

Another explanation for the greater verifiable impact of formative assessments on the top two-thirds of the students (measured by LSAT scores and UGPA) is that these students may focus more on grades or scores as a critical measure of success, and therefore are more thorough and diligent in using the materials to maximize scores. College students typically adopt surface, deep, or strategic approaches to learning and these approaches can impact their academic outcomes.<sup>80</sup> The main goal of deep learners is to learn and understand; surface learners complete required tasks but without interest in learning; and strategic learners attempt to get high grades, avoiding activities that jeopardize scores and maximizing activities that improve scores. A study using law students found that those who focused on achieving high marks had higher LSAT scores.<sup>81</sup> Accordingly, those with higher LSAT scores

77. Anthony S. Niedwiecki, *Lawyers and Learning: A Metacognitive Approach to Legal Education*, 13 *Widener L. Rev.* 33, 35 (2006) (noting that “[g]enerally metacognition refers to having both awareness and control over one’s learning and thinking”); John L. Nietfeld, Li Cao & Jason W. Osborne, *Metacognitive Monitoring Accuracy and Student Performance in the Postsecondary Classroom*, 74 *J. Experimental Educ.* 7, 9 (2005) (noting that metacognition “help[s] learners use their attentional resources more efficiently, process information at a deeper level, and monitor their performance more accurately”).
78. *See Does Practice Make Perfect*, *supra* note 11, at 280–82 (discussing interplay between metacognition and student learning and performance). Sadler, *supra* note 12, at 121 (noting that feedback can be used to narrow the achievement gap if students understand the standards, compare their performance with the standard and engage in action to close the gap between the standard and their performance).
79. Srinivasan, *supra* note 47, at 862–63.
80. Marann Byrne, Barbara Flood & Pauline Willis, *An Inter-Institutional Exploration of the Learning Approaches of Students Studying Accounting*, 20 *Int’l J. Teaching & Learning in Higher Educ.* 155, 156 (2009).
81. Leah M. Christensen, *Enhancing Law School Success: A Study of Goal Orientations, Academic Achievement and the Declining Self-Efficacy of Our Law Students*, 33 *Law & Psychol. Rev.* 57, 70–71 (2009).

may be more willing or able to “slavishly copy the exemplars”<sup>82</sup> to achieve their grade goals than those who benefitted less from the materials.<sup>83</sup>

Our results may also reflect that diligent use of formative assessments is dependent on a student’s sense of confidence that the materials will help them. Most law students never get feedback after a final exam, and few review their final exam answers. So, the reasons underlying their grades often remain a mystery.<sup>84</sup> Reducing uncertainty about how one can achieve good grades may lead to higher motivation, more efficient studying strategies,<sup>85</sup> and greater confidence that studying harder will produce better grades. The motivational psychology literature contains decades of work with college students showing that higher confidence leads to increased academic effort and resiliency.<sup>86</sup> Seeing what was expected and having the chance to practice their skills may have increased some students’ confidence that they could meet the challenges of learning and apply the substantive law, and so they intensified their efforts to do so.<sup>87</sup>

### Limitations of the Study

The unequal class time between comparison groups may have suppressed some of the learning effects. The formative assessments were completed and/or reviewed during class, taking about five total hours, so the control group had more class hours to spend on course topics. In other words, the feedback

82. Sadler, *supra* note 12, at 128 (noting that giving students examples of high quality work runs the risk of students copying the example without understanding the reasons why the example constituted high quality work).
83. See, e.g., P. Ramsden, D.G. Beswick & J.A. Bowden, Effects of Learning Skills Interventions on First Year University Students’ Learning, 5 *Hum. Learning* 156 (1986) (discussing a study in which learning skills programs designed to deepen student learning actually resulted in students adopting a more surface learning approach in order to achieve short-term grade goals).
84. Educating Lawyers, *supra* note 1, at 165.
85. Shute, *supra* note 15, at 157 (discussing studies which show that uncertainty is “an aversive state that motivates strategies aimed at reducing or managing it” and suggesting that “reducing uncertainty may lead to higher motivation and more efficient task strategies”).
86. Albert Bandura, Perceived Self-Efficacy In Cognitive Development and Functioning, 28 *Educ. Psychologist* 117 (1993); Reid Bates & Samer Khasawneh, Self-Efficacy and College Students’ Perceptions and Use of Online Learning System, 23 *Computers in Hum. Behav.* 175 (2007); Gian Vittorio Caprara et al., Longitudinal Analysis of the Role of Perceived Self-Efficacy for Self-Regulated Learning in Academic Continuance and Achievement, 100 *J. Educ. Psychol.* 525 (2008); *but see* Christensen, *supra* note 81, at 71–73 (finding that the law students with the highest grades were often the least confident).
87. Albert Bandura & Daniel Cervone, Differential Engagement of Self-Reactive Influences in Cognitive Motivation, 38 *Organizational Behav. & Hum. Decision Processes* 92, 108 (1986); *see also* Kluger, *supra* note 21, at 260 (noting that students are more likely to increase effort when the intended goal is clear).

didn't just have to be helpful, it had to be more helpful than additional class time. In a future study, both groups could have equal class time if the materials were provided as handouts or posted in course management software. This suggestion comes with a caveat: When students see the answer before attempting the problem, student achievement is diminished rather than enhanced.<sup>88</sup> Therefore, the model answer should come after students have completed the problem.<sup>89</sup> Adjusting the implementation in this way may also alter the results because making students retrieve materials on their own time requires higher motivation than completing them in class.

Although this study indicates that formative assessments improved student performance, in particular for those with higher LSAT/UGPAs, there may be a Hawthorne effect, *e.g.* students did better because they knew that their performance was being studied.<sup>90</sup> The second author created the study materials after the 2008 course was complete so only the 2009 students were aware that their performance was being monitored. Thus, the 2009 students' performance may have been impacted by their desire to please the investigator who was also their professor. However, if that were true, one might expect an across-the-board increase in performance rather than a stronger effect for those with higher LSAT scores/UGPAs.

Given the large portion of the final exam that was identical in 2008 and 2009, students might somehow have heard about the content of the questions from colleagues who had taken this course in the prior year and thus been able to better prepare for the exam. Most 2008 students did not review the exam following receipt of their grade and those who did reviewed the exam on school premises and left it with the instructor. Additionally, if exam questions were leaked, we would expect an across the board improvement rather than seeing the performance improvement focused in the top two-thirds of LSAT score or UGPA students.

The quiz and final exam questions in this study were generally single-issue questions involving short answers or short essay responses in which there was generally a correct response. The gains made by students in this class may not have been as strong or across such a wide range of students in a course in which the substantive material was tested with a more traditional essay exam. However, developing students' ability to spot issues and analyze facts to determine the applicable rule is a stepping stone to answering more analytically

88. Bangert-Drowns, *supra* note 15, at 224, 233.

89. For instance, course management software can require students to respond to questions before the model answer is displayed.

90. Frank Merrett, Reflections on the Hawthorne Effect, 26 *Educ. Psychol.* 143, 146 (2006) (noting that studies cannot look merely at outcomes but also must consider factors such as the information accompanying the experiment and that information's effect on study participants).

complex multi-issue questions. Thus, the value of formative assessments in improving students' performance should not be discounted simply because of the exam format.

Looking at this study's results in light of Christensen's study about mastery versus performance-oriented learners<sup>91</sup> raises questions about whether practice materials inadvertently encourage performance-oriented goals, rather than encouraging deeper mastery learning. In other words, do practice materials support those whose main goal is to get higher course grades rather than assisting those who wish to truly comprehend and master the content? While these goals are not necessarily mutually exclusive, professors typically prefer to downplay grade goals in favor of mastery goals. This leaves a lingering question about whether formative assessments are a positive addition to law school assessment culture. Fortunately there is significant work in the literature suggesting that the nature of exam questions strongly influence student study practices and learning approaches,<sup>92</sup> so if the practice exam questions call for deep learning, students will adopt deep approaches to learning.

This study had a relatively small sample size, especially when sub-dividing participants into the top two-thirds and bottom one-third of the class by certain measures, so the results should be treated with caution. The lack of effect for one-third of the students in the intervention class could be a function of low statistical power in the regression model, although visual inspection of scores (such as Figure 1) confirm the statistical results.

Students completed much of the formative assessment material in class so even those with low interest in the materials complied with the instructor's request to answer the questions. However, they may not have had genuine interest in using the materials to change their study patterns. In future work, loading the materials into a course management software (such as WebCT) and tracking the number of questions completed, model answers reviewed, and reflective tasks completed, would permit tracking the motivation to use formative assessments. This might shed light on whether the materials were used but not effective, or largely ignored by those who did not benefit.

Some educators distinguish between formative assessments that are one-size-fits-all materials where everyone in the class gets the same material, as in this work, and activities that are tailored to the learner's prior knowledge and

91. See Christensen, *supra* note 83.

92. Terence J. Crooks, The Impact of Classroom Evaluation Practices on Students, 58 Rev. Educ. Res. 438, 444-47 (1988) (reviewing studies demonstrating that students adapt their study approaches to what the exam structure requires, thus if exams call for surface learning, students use a surface approach, but if the exam calls for deeper learning, students use a deeper learning approach).

individual ability.<sup>93</sup> Perhaps the students who did not benefit would respond to materials better adapted to their learning approach, prior knowledge, and individual ability. While it is impractical to tailor materials to individual students, it is also unclear whether minor tailoring such as offering a starter set at course onset would expand the benefits to students getting undetectable benefits with the existing materials.

### **Ease of Implementing Formative Assessments**

While drafting the questions, model answers, rubrics, and self-reflective exercises initially takes a few hours, those materials do not need updating each term. Grading a short midterm also takes a few hours, but may result in faster final exam grading due to better quality responses. Alternately, giving a midterm may justify a shorter final exam, thereby reducing time spent grading final exams.

In a meta-analysis of three decades of work on feedback to college students, one key to improved performance was an explanation of why an answer was correct or incorrect.<sup>94</sup> This kind of feedback is possible without individually grading each student answer. Therefore, it may be possible to produce the same learning effect without having to grade a mid-term, further minimizing instructor effort. A future study duplicating the model described here but eliminating the instructor graded mid-term could shed light on the question of whether the same effect could be found when providing model answer (ungraded) midterms.

### **Conclusion**

In this study, the final exam scores of students in the intervention group on eleven common questions increased from six to nine percentage points for about 70 percent of the class as the result of providing five ungraded quizzes, a graded midterm, model answers, grading rubrics and self-reflection activities during the semester. The improvement was seen in some students with below-the-median first-year law school grades, as long as those students were in the top two-thirds of the class on UGPA or LSAT scores.

The results add to the previous study,<sup>95</sup> reinforcing the power of formative assessments in law classes. The data show that formative assessments can improve students' final exam scores for a majority of students, and that some students with weak first year grades may catch up to their peers with feedback. However, the benefit seems to accrue disproportionately to students who are in the top two-thirds in terms of LSAT/UGPA, perhaps due to their desire or ability to adjust to feedback, their higher confidence in their own ability to effectively use feedback, and their ability to better self-monitor and calibrate their comprehension.

93. Rushton, *supra* note 7, at 511.

94. Bangert-Drowns, *supra* note 15, at 232.

95. Does Practice Make Perfect, *supra* note 11.

While students with both above and below the median first-year law school grades improved, our tests did not detect learning advantages for one-third of the students (those with the lowest LSAT scores and UGPAs). It is unclear whether additional practice, more extensive reflective exercises, or different kinds of practice could provide benefits to that last one-third of the class.

In sum, we have provided new evidence that shifting the law school culture away from a single summative assessment may advantage students. We have also proposed how to do this work without unreasonably burdening faculty. We believe this work highlights a win-win that should advance the wide-scale experimentation and adoption of good formative assessment practices in law classes.

## Appendix A

### *SAMPLE QUIZ & MIDTERM QUESTIONS*

#### *Sample Quiz Questions*

Scenario:

Don is on trial for criminal assault and battery. He is accused of beating up Adam at a local bar. Adam suffered a broken nose, various bruises and also lost two teeth when Don punched him in the face. Don's defense is that he acted in self-defense.

Don claims that he accidentally bumped Adam and poured beer on him. He says that even though he tried to apologize, Adam began screaming at him, shoved him and began swinging. At that point, Don punched Adam a few times in the face because Don was afraid that Adam was about to become violent and Don needed to stop him. Adam has a different version of what happened. Adam claims that he turned to Don after the beer was spilled and said something like, "Hey Buddy, watch what you're doing" and then was turning back to talk to his friends when Don started punching him. Don and Adam went to the same college and knew of each other but had had no social interactions before this fight. No one else at the bar saw exactly what happened until Don began hitting Adam. Don would like to introduce the following testimony on his direct examination.

Question: "What, if anything, had you heard about Adam's temper, before the night of the bar fight?"

Answer: "I heard that a few years ago, Adam got really drunk and took a knife and slashed the face of some guy who was flirting with his girlfriend."

The State objects to the introduction of this evidence. In this jurisdiction, the law on self-defense is: the actor uses reasonable force to defend against unprivileged acts and the actor has a reasonable belief force is necessary to prevent harmful or offensive contact.



*Student Requirements*

1. Using ONLY the rules we have studied thus far, make all possible objection[s] and accompanying arguments in support of the objection[s] on behalf of the State. Make Don's viable responses to those arguments and decide how the judge should rule.

HINT: Ask yourself—"what is the evidence being offered to prove?"

2. If the judge excludes the evidence, what must Don do if he wants to preserve the issue for appeal and how should he go about doing it?

*Sample Midterm Questions*

Scenario:

Al Jones has been charged with conspiring to bomb the Federal Courthouse in Atlanta. At trial, the government calls Cindy Adden, an FBI agent to the stand to testify as a lay witness. Here's the relevant transcript excerpt.

Question: Please state your name and occupation.

Answer: My name is Cindy Adden. I am a special investigator in the domestic terrorism unit of the F.B.I.

Question: How did you come to be involved in this case?

Answer: Through a network of underground informers, I heard about a potential plot to bomb the Federal Courthouse in Atlanta within the next couple of days. Based upon that information, I got a search warrant for the defendant's apartment. OBJECTION and MOTION TO STRIKE.

*Student Requirement:*

Identify the proper objection [note that you are NOT to discuss hearsay issues] and in 125 words or less, make the defendant's argument. [3 points—13 minutes]

*Continuation of scenario:*

The following question involves evidence presented by the defendant. The defendant calls his neighbor, Sandy Sims, to the stand. Sims testifies as follows:

Sims: I have gotten to know Mr. Jones very well in the past couple of years. He has been incredibly kind and helpful. He has watched my children when I went for a job interview, lent me money when I had to take my child to the doctor and couldn't afford it and even driven me to work when I first got a job and didn't have money to get my car fixed. He is a kind and gentle man and not the type of person who would bomb a building. OBJECTION AND MOTION TO STRIKE.

*Student Requirement:*

Identify the government’s objection to this testimony and in 125 words or less make the government’s argument in support of its objection. [3 points—13 minutes]

*Continuation of scenario:*

During the defendant’s case, the defendant testified that the computer the FBI found belonged to his friend, Allen Redkin. In its rebuttal case, the government calls Allen Redkin. Redkin pled to a lesser charge and is awaiting sentencing.

Question: Mr. Redkin, did the laptop computer found in defendant’s apartment belong to you?

Answer by Redkin: Yes, it was mine.

After this testimony, the prosecutor approaches the bench and says to the judge, “your honor, during plea negotiations, Mr. Redkin said that the computer belonged to the defendant. I’d like permission to impeach him on this critical fact.”

*Student Requirement:*

As the judge, make a ruling on this request and in 100 words or less explain your reasoning. [2 points—9 minutes]

**Appendix B***SAMPLE SELF-REFLECTIVE QUESTIONS**Beginning of Semester Questions (general in nature):*

1. Overall, did I identify all issues [relevance; 403 probative value & risk of prej; offer of proof—how you make it and what you must show]? If I missed some issues, what can I do to improve my issue spotting?
2. Did I argue both sides for each issue? If I missed an argument, how can I help myself remember to try and address both sides of each issue?
3. Did I use all relevant facts for each issue [i.e. did I say—it’s not relevant because it was a long time ago and involved flirting; did I say—unfairly prejudicial because jury will judge Adam as a bad actor]?—Key here—did I use the word “because” and then use the facts or an inference from the facts to explain my statement? If not, what can I do to help train myself to do this?

*Middle of Semester Questions (more specific):*

1. This question required you to use 404(b) (prior bad acts) as the basis of your analysis. Did you identify that rule? What words in this problem clue you

in to the fact that you have to look to 404(b)? Identify, by underlining, where in your answer you stated that the problem involved rule 404(b).

2. Did you identify the specific non-propensity purpose for which this evidence would be offered [i.e. knowledge]? What words in this problem clue you in to the fact that you have to offer it to prove knowledge?

3. Did you explain how this evidence would go to the defendant's knowledge that Dylan would take the drugs (and explain HOW it went to that knowledge)? Specifically, what facts did you use to support your argument that it went to her knowledge that Dylan would take the drugs? Were there any others you could have used? Identify, by underlining, where in your answer you discussed how the evidence was being offered to prove Dylan's knowledge.

4. Did you get off track by trying to make this a "character in issue" case? If so, why did you make that argument? Do you understand why this is not a case of "character in issue" but instead a 404(b) analysis?

5. Did you only make ONE argument [the question asks for the BEST argument]?

*Accompanying Graded Mid-term (shift in format and approach)*

Using the grading rubric provided for each question, identify the points you got [identify on your answer exactly where each point comes from—e.g. if it says "explain how the probative value is low—identify where in your answer you explained how the probative value was low].

Overall, looking at your answer in light of the rubric, which of the following do you think you should be working on between now and the final [check all that are applicable]:

- \_\_\_\_\_ making sure I understand the elements of the rule
- \_\_\_\_\_ making sure I understand the exceptions to the rule
- \_\_\_\_\_ practice issue spotting
- \_\_\_\_\_ practice articulating clearly and concisely how the rule applies to the facts
- \_\_\_\_\_ other—describe

Which of the following practice study tips do you think will help you learn and apply this material [check all that are applicable]:

- \_\_\_\_\_ doing an outline of the rules and exceptions
- \_\_\_\_\_ using Emanuel's or Cali exercises
- \_\_\_\_\_ writing out answers to Emanuel's or Cali Exercises and exchanging those with a colleague
- \_\_\_\_\_ other—describe

Comparing your own scoring to the score you got from the professor, were the scores similar? If not, when did you give yourself points that the professor did not give you? Is there any pattern that you can see about when you scored yourself higher or lower?