



**THE OHIO STATE UNIVERSITY**  
MORITZ COLLEGE OF LAW

**Formative Assessments: A Law School Case Study**

Deborah Jones Merritt  
Ruth Colker, Ellen Deason, Monte Smith, Abigail Shoben

**Public Law and Legal Theory Working  
Paper Series  
No. 392**

**April 19, 2017**



This working paper series is co-sponsored by the  
Center for Interdisciplinary Law and Policy Studies  
at the Moritz College of Law

This paper can be downloaded without charge from the  
Social Science Research Network:  
<http://ssrn.com/abstract=2955055>

## FORMATIVE ASSESSMENTS: A LAW SCHOOL CASE STUDY\*

Ruth Colker,<sup>\*\*</sup> Ellen Deason,<sup>\*\*\*</sup> Deborah Merritt,<sup>\*\*\*\*</sup> Abigail Shoben<sup>\*\*\*\*\*</sup> & Monte  
Smith<sup>\*\*\*\*\*</sup>

### I. INTRODUCTION

American Bar Association (“ABA”) Standard 314, which took effect during the 2016-17 academic year, says: “A law school shall utilize both formative and summative assessment methods in its curriculum to measure and improve student learning and provide meaningful feedback to students.”<sup>1</sup> Summative assessments are tests given at the end of the semester, which seek to measure how much students have learned; formative assessments also measure student learning but offer feedback so that students can improve their performance in the future.<sup>2</sup> Traditionally, law schools provided nearly exclusively summative assessments outside of writing or clinical courses. The new ABA

---

\* This project has received IRB approval, no. 2016E0360. No identifiable student information is available in our data set.

\*\* Distinguished University Professor & Heck-Faust Memorial Chair in Constitutional Law, Moritz College of Law, The Ohio State University.

\*\*\* Joanne Wharton Murphy/Classes of 1965 and 1973 Professor of Law, Moritz College of Law, The Ohio State University

\*\*\*\* John Deaver Drinko-Baker & Hostetler Chair in Law, Moritz College of Law, The Ohio State University.

\*\*\*\*\* Associate Professor, College of Public Health, The Ohio State University.

\*\*\*\*\* Assistant Dean for Academic Affairs, Moritz College of Law, The Ohio State University.

<sup>1</sup> See American Bar Association, *Managing Director’s Guidance Memo*, Standards 301, 302, 314 and 315, June 2015,

[http://www.americanbar.org/content/dam/aba/administrative/legal\\_education\\_and\\_admissions\\_to\\_the\\_bar/governancedocuments/2015\\_learning\\_outcomes\\_guidance.authcheckdam.pdf](http://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/governancedocuments/2015_learning_outcomes_guidance.authcheckdam.pdf).

<sup>2</sup> See Carol Springer Sargent & Andrea A. Curcio, *Empirical Evidence that Formative Assessments Improve Final Exams*, 61 J. LEGAL EDUC. 379, 381 (2012) (“Formative assessments seek to increase learning and motivation by offering students feedback about gaps between current and desired levels of performance. Summative assessments, by contrast, seek to measure the amount of learning.”); Andrea A. Curcio et al., *Essay Question Formative Assessments in Large Section Courses: Two Studies Illustrating Easy and Effective Use* in EXPLORING LEARNING & TEACHING IN HIGHER EDUCATION 349-368(2014) (“Formative assessments measure learning for the purpose of giving feedback rather than for the purpose of assigning a grade.”)

standard reflects the emerging literature that formative assessments assist student learning, especially for students with a growth mindset who might adjust their learning styles or study habits in response to feedback.<sup>3</sup> Although ABA Standard 314 only requires law schools to document that they are making formative assessment opportunities available to students, the actual goal is likely to make *effective* formative assessments available to students.

This article is the first of what we hope will be several articles exploring the results from natural experiments at the Moritz College of Law at The Ohio State University regarding what kinds of formative assessments might best enhance student learning. “Natural experiments” refer to comparisons based on differences in students’ experiences that arise due to class scheduling and professors’ practices, but are not planned ahead of time as a formal experiment.<sup>4</sup>

This article discusses empirical results from a natural experiment in Professor Ruth Colker’s required Constitutional Law class, which was taught to one-third of the first-year class during the spring semester in 2014, 2015, and 2016. Professor Colker offered students the opportunity to compose an answer to one question from a previous year’s final examination, submit the answer to her, and receive both an estimated numerical grade and detailed feedback on the answer. The estimated grade did not factor

---

<sup>3</sup> “[F]eedback effectiveness turns not just on the materials provided, but also on the ability of the recipient to digest and use the feedback, as well as their goals, self-confidence, interest, and intentions.” Sargent & Curcio, *supra* note 2, at 383.

<sup>4</sup> Other authors have recently made use of natural experiments based on law school scheduling. See, e.g., Daniel E. Ho & Mark G. Kelman, *Does Class Size Affect the Gender Gap? A Natural Experiment in Law*, 43 J. LEGAL STUDS. 291 (2014); Daniel Schwarcz & Dion Farganis, *The Impact of Individualized Feedback on Law Student Performance*, 67 J. LEGAL EDUC. (forthcoming 2016), [https://papers.ssrn.com/sol3/papers2.cfm?abstract\\_id=2772393](https://papers.ssrn.com/sol3/papers2.cfm?abstract_id=2772393).

into the students' final grade in the course. Instead, the exercise provided purely formative feedback in the hope of enhancing student learning.

This exercise was optional; about half the students participated. The voluntary nature of the exercise allows us, with appropriate controls, to assess the relationship between this type of formative feedback and student performance. We examined three questions: (1) Were some students more likely than others to seek this formative feedback? (2) Was receipt of this feedback associated with higher grades on the course's final exam, after controlling for other factors that might predict performance? (3) Was the receipt of formative feedback in one class associated with better performance in the student's *other* spring-semester classes, after controlling for similar factors?<sup>5</sup>

Part II of this article will review the literature on formative feedback and discuss our study's design. Part III will report the results of our natural experiment. Part IV will reflect on the meaning of those results and place them in the context of other studies of formative feedback. Although we can only show a correlation between receiving one type of formative feedback and student learning, we conclude that the combination of our study with other work in the area suggests that a causal relationship may exist between at least some formative feedback and enhanced student learning.

## II. BACKGROUND

### A. Literature Review

Many articles discuss the importance of formative feedback to student learning, and the substantial need for legal education to be improved through increased use of

---

<sup>5</sup> We were stimulated to ask this third question by a study conducted at the University of Minnesota Law School, which suggested that students who received formative feedback during the semester outperformed their peers (who did not receive this feedback) in another concurrent or subsequent class. *See* Schwarcz & Farganis, *supra* note 4, at 20 tbl.1, 22-23, 23 fig.3.

effective formative feedback.<sup>6</sup> Although numerous scholars have conducted empirical assessments to evaluate the relationship between formative feedback and academic achievement with high school and college students,<sup>7</sup> relatively few researchers have attempted to evaluate this relationship in the law school setting.<sup>8</sup> In one of the earliest attempts, in 1981, Professor Gary A. Negin devised an empirical study to determine whether frequent testing and feedback in a first-year Torts class would correlate with increased academic achievement,<sup>9</sup> consistent with findings that frequent undergraduate testing was associated with improved academic achievement.<sup>10</sup>

In his law school empirical study, Professor Negin randomly divided<sup>11</sup> 75 first-year law students in a Torts class into three groups.<sup>12</sup> Group 1 received four multiple-choice examinations, three given at monthly intervals and a fourth as a final exam.<sup>13</sup> Group 2 received two examinations, a midterm given at the same time that Group 1 received its second examination and combining the questions from Group 1's first two

---

<sup>6</sup> For a recent thorough review of this literature, see Schwarcz & Farganis, *supra* note 4, at 7-10.

<sup>7</sup> See, e.g., William R. Balch, *Practice Versus Review Exams and Final Exam Performance*, 25 TEACHING PSYCHOL. 181, 182-83 (1998) (college-level introductory psychology students); John A. Gretes & Michael Green, *Improving Undergraduate Learning with Computer-Assisted Assessment*, 33 J. RESES. ON COMPUTING EDUC. 46, 48 (2000) (undergraduates in an education course); James A. Kulik et al., *Effects of Practice on Aptitude and Achievement Test Scores*, 21 AM. EDUC. RES. J. 435 (1984) (meta-analysis of studies involving high school and college students).

<sup>8</sup> See, e.g., Andrea A. Curcio et al., *Developing an Empirical Model to Test Whether Required Writing Exercises or Other Changes in Large-Section Law Class Teaching Methodologies Result in Improved Exam Performance*, 57 J. LEGAL EDUC. 195 (2007); Andrea A. Curcio et al., *Does Practice Make Perfect?: An Empirical Examination of the Impact of Practice Essays on Essay Exam Performance*, 35 FLA. ST. U. L. REV. 271 (2008) [hereinafter *Does Practice Make Perfect?*]; Andrea A. Curcio et al., *Essay Question Formative Assessment in Large Section Courses: Two Studies Illustrating Easy and Effective Use in EXPLORING LEARNING & TEACHING IN HIGHER EDUC.* 349-368 (2014); Sargent & Curcio, *supra* note 2, at 381.

<sup>9</sup> Gary A. Negin, *The Effects of Test Frequency in a First-Year Torts Course*, 31 J. LEGAL EDUC. 673 (1981).

<sup>10</sup> *Id.* at 673-74.

<sup>11</sup> *Id.* Professor Negin did not assign students to groups on the basis of LSAT scores, undergraduate grade point averages ("UGPA"), or membership in particular demographic groups, and his study did not attempt to control for any of those factors.

<sup>12</sup> *Id.* at 674.

<sup>13</sup> *Id.* at 675.

examinations, and the same final examination that Group 1 received.<sup>14</sup> Group 3 received only the same final examination that Groups 1 and 2 received.<sup>15</sup> Professor Negin compared the performances of Groups 1 and 2 in the middle of the semester and of all three groups on the final examination.<sup>16</sup>

Professor Negin found a significant relationship between formative feedback and academic performance, with the top-performing group having received the most feedback.<sup>17</sup> He observed that a statistically significant gap between Groups 1 and 2 had appeared by the point in the semester at which Group 1 had received its second 75-question examination and Group 2 had received its first 150-question examination.<sup>18</sup> Group 1 maintained, but did not increase, that gap on the final examination.<sup>19</sup> Professor Negin concluded, therefore, that “academic achievement could be improved if more than one test was given.”<sup>20</sup>

Little in the way of empirical studies of the relationship between formative assessment and law student academic achievement appears in the literature between Professor Negin’s 1981 report and 2004.<sup>21</sup> In 2008 and 2014, Professor Andrea A. Curcio

---

<sup>14</sup> *Id.*

<sup>15</sup> *Id.*

<sup>16</sup> *Id.* at 676.

<sup>17</sup> Out of a possible 225 points, Group 1 scored 1985.52, Group 2 scored 192.72 and Group 3 scored 186.72. These results were statistically significant at the .025 level. *Id.* at 676, tbl. 2.

<sup>18</sup> These results were statistically significant at the .05 level. *Id.* at 676.

<sup>19</sup> *Id.*

<sup>20</sup> *Id.*

<sup>21</sup> Professor Charles A. Rees conducted one such study in 2004. *See* Charles A. Rees, *The ‘Non-Assessment’ Assessment Project*, 57 J. LEGAL EDUC. 521 (2007). The study involved students in two different classes taught by Rees, one in which he gave multiple formative assessments, both formal and informal, and one in which he gave none. He compared the performances of the two groups on similar assessments and concluded that the students who had received extensive feedback performed at the same level as the students who had received minimal feedback. *Id.* at 523. The two classes were Constitutional Law and Civil Procedure II, and Professor Rees did not control for differences between the two classes, either in terms of the academic credentials of the students enrolled or in subject matter. He acknowledges that he was an “old teacher,” in terms of years of experience teaching the subject matter, in one class and a

and her colleagues reported two empirical studies on the relationship between individualized feedback and law student achievement,<sup>22</sup> and, in 2016, Professor Daniel Schwarcz and his student, Dion Farganis, published a study on the relationship between individualized feedback and law student performance.<sup>23</sup> The results of their studies influenced the development of our empirical project.

In her first study, Professor Curcio and her colleagues evaluated the relationship between a combination of generalized and individualized feedback on writing assignments and students' grades on a final examination essay in Civil Procedure. Students in one Civil Procedure class wrote five "practice papers" calling for analyses of increasingly challenging issues during the semester.<sup>24</sup> The professor primarily provided generalized feedback on the exercises in the form of annotated model answers.<sup>25</sup> Students used the model answers for self-assessment or (for one assignment) peer-assessment during class time.<sup>26</sup> The professor also provided individualized feedback to each student in the form of a grade and comments on one paper.<sup>27</sup> Students in another professor's Civil Procedure class did not write the practice papers.<sup>28</sup>

The two professors collaborated on a final essay examination and concluded that performance by the students in the class with writing exercises and feedback was

---

"new teacher" in the other. *Id.* at 522. The differences between the two classes make it difficult to interpret the effect of the different feedback.

<sup>22</sup> See *Does Practice Make Perfect?*, *supra* note 8; Sargent & Curcio, *supra* note 2.

<sup>23</sup> See Schwarcz & Farganis, *supra* note 4. Before attending law school, Farganis was an Assistant Professor of Political Science at Elon University.

<sup>24</sup> *Does Practice Make Perfect?*, *supra* note 8, at 287. Each assignment was a single-issue essay question involving a legal rule that had been discussed in class. *Id.*

<sup>25</sup> *Id.* at 287-88. For the first assignment, the professor also provided feedback through a class discussion of common problems. For one paper, the professor also provided a grading rubric. *Id.*

<sup>26</sup> *Id.* at 288.

<sup>27</sup> *Id.* at 287-88. Students were also given the opportunity to meet one-on-one with an intervention specialist, although few availed themselves of the opportunity. *Id.* at 288-89.

<sup>28</sup> See *id.* at 286-87.

significantly better as assessed by both professors through anonymous grading.<sup>29</sup> Notably, Professor Curcio and her colleagues found that “students who had received the writing interventions had higher average raw scores on each of the [final exam] essay questions.”<sup>30</sup> The stronger performances were not, however, distributed evenly throughout the class; the deviation in performance for students who received the feedback was most significant for students with above-median LSAT scores and UGPAs.<sup>31</sup> In addition, the relationship between doing the exercises and improved academic performance did not extend beyond the Civil Procedure class; there was no significant relationship between completing these exercises and academic performance in other classes.<sup>32</sup>

Professor Curcio acknowledged that the study was subject to criticism for lack of control over, among other factors, differences in teaching experience and style between the two professors and ability of students assigned to the two sections of the class.<sup>33</sup> In a second study, published in 2014, Professor Curcio, working with Professor Carol Springer Sargent, eliminated the effect of different professors by testing the effects of feedback on two groups of students taught the same subject matter by the same professor in successive years.<sup>34</sup>

---

<sup>29</sup> See *id.* at 290-92.

<sup>30</sup> *Id.* at 299. Students who had received the writing interventions earned average scores of 19.555 on the first essay question and 15.09 of the second. *Id.* at 291. Students who had not received the interventions earned average scores of 17.18 and 12.135, respectively on the two questions. *Id.*

<sup>31</sup> *Id.* at 294-97.

<sup>32</sup> *Id.* at 298-99.

<sup>33</sup> *Id.* at 291.

<sup>34</sup> See Sargent & Curcio, *supra* note 2, at 384.



Professor Curcio taught Evidence to two groups of students, one in 2008 and the other in 2009.<sup>35</sup> The 2008 class was the control group, and Professor Curcio taught that group using “a problem method supplemented by case analysis.”<sup>36</sup> She assessed students’ performance using only one cumulative final exam.<sup>37</sup> In 2009, Professor Curcio taught using the same methods but added a series of formative assessments consisting of five ungraded quizzes and a graded midterm.<sup>38</sup> She primarily provided feedback through model answers, grading rubrics, and self-assessment exercises.<sup>39</sup> Students in the 2009 class took a final examination using some identical questions to those in the 2008 exam.<sup>40</sup> Performance on these common questions was used to compare the two years.<sup>41</sup>

In Professor Curcio’s Evidence class study, the second group of students (who had received interim feedback) significantly outperformed the control group. The difference averaged to 3.024 points out of 50, which was equal to about a half letter grade or 6.048 percent.<sup>42</sup> This study also attempted to assess the relationship between the effectiveness of the feedback and measures of prior student achievement. The positive effect of the feedback was present among students with LSAT scores at or above the school median, but not for those with below-median scores.<sup>43</sup> Professor Curcio did not

---

<sup>35</sup> Evidence was a required second-year course at Professor Curcio’s school. *Id.* at 385. This study differs from the others reported in this section, as well as from our own study, because it assessed second-year students rather than first-year ones.

<sup>36</sup> *Id.*

<sup>37</sup> *Id.*

<sup>38</sup> *Id.*

<sup>39</sup> *Id.* at 385, n.44. In addition, the graded midterm included comments from the professor focused on suggestions for improvement. *Id.* at 387 & n.54.

<sup>40</sup> *Id.* at 389.

<sup>41</sup> *Id.* Scores on the common questions were highly correlated to the overall scores on the final exam. *Id.*

<sup>42</sup> *Id.* at 391. This regression analysis controlled for UGPA, LSAT, and first-year law grades. The results were statistically significant at  $p < .05$ .

<sup>43</sup> *Id.* Similarly, the coefficient for feedback was significant in a regression equation that included students with the top two-thirds of the class as measured by UGPA; the coefficient was not significant for students in the bottom third of the class by that measure. *Id.* at 391-92. The small size of the latter group, however,

report on the effect, if any, in the second group's performance in classes other than Evidence. In neither study did Professor Curcio test differences among demographic groups.

Professor Schwarcz and his co-author Farganis attempted to ameliorate some of the limitations in Professor Curcio's studies by capitalizing on a natural experiment resulting from the occasional class that combined two sections in a first-year law school cohort into a "double section" at the University of Minnesota Law School.<sup>44</sup> These double-section classes allowed Schwarcz and Farganis to compare the performance of students who had received individualized formative feedback in a small section to that of students within the double section who had not. Because the study examined multiple courses taught by different professors, it included many forms of feedback. The study was limited to feedback on examinations, or on assignments with issue-spotter or policy questions similar to those a student would experience on an examination.<sup>45</sup> Individualized feedback was defined to include assigning grades on exams, providing written comments to individual students, and providing oral comments in small groups or to students individually.<sup>46</sup> Thus, the analysis combined diverse graded and ungraded feedback experiences ranging from a grade on a multiple-choice examination to oral commentary on practice "issue spotter" questions.<sup>47</sup>

The natural experiment Schwarcz and Farganis analyzed builds on the previous studies in significant ways. First, the study had natural control groups in the form of

---

makes interpretation of this finding difficult; coefficients are less likely to reach significance in a smaller sample.

<sup>44</sup> See Schwarcz & Farganis, *supra* note 4, at 4.

<sup>45</sup> *Id.* at 16-17. The study excluded formative feedback in legal writing courses. *Id.*

<sup>46</sup> *Id.* at 17-18, 18 fig.1. Model answers, grading rubrics, and generalized oral feedback were not treated as individualized feedback for purposes of this study. *Id.* at 17.

<sup>47</sup> *Id.* at 17-18, 18 fig.1, 20-21 & n.56.

sections that were not provided individualized feedback but were paired in double sections with classes that were provided such feedback.<sup>48</sup> Second, because Schwarcz and Farganis devised the study after the examined classes had ended, none of the professors could have been affected by their knowledge that the students were part of a study.<sup>49</sup> Finally, Schwarcz and Farganis had data from all classes in which the students in the study received grades, so they were able to test whether there was a relationship between receiving individualized feedback and performance in other classes.<sup>50</sup> The design of the natural study does, however, have limitations in that different professors taught the comparison groups and feedback came in many different forms. As Schwarcz and Farganis recognize, the professors who provided feedback might have been effective for other reasons that also promoted higher student performance.<sup>51</sup>

Schwarcz and Farganis found mean grade differences in favor of the students in the segment of each double section that received individualized feedback. The difference did not reach statistical significance within any one double section,<sup>52</sup> but was significant for all sections combined.<sup>53</sup> A regression analysis that controlled for other factors commonly associated with first-year grades, moreover, confirmed that receiving feedback was significantly associated with higher course grades.<sup>54</sup>

---

<sup>48</sup> *Id.* at 15, 19-20, 20 tbl.1.

<sup>49</sup> *Id.* at 15-16.

<sup>50</sup> *Id.* at 16. Professor Curcio tested this in her first study, but did not find significant effects. *Does Practice Make Perfect?*, *supra* note 8, at 306-07.

<sup>51</sup> Schwarcz & Farganis, *supra* note 4, at 33-34. Schwarcz and Farganis examined this possibility by examining student evaluations of the clarity of their professors. This variable did not make a statistically significant contribution to the prediction of student grades, supporting a conclusion that it was not a substantial factor. *Id.* at 34-36.

<sup>52</sup> *Id.* at 23 & fig.3.

<sup>53</sup> *Id.* at 23-24.

<sup>54</sup> *Id.* at 27-28, 28 tbl.2.

The associations were not, however, distributed evenly. Schwarcz and Farganis found that the correlation between receiving feedback and student grades was “most pronounced” for the students whose grades were below the mean performance in their double section.<sup>55</sup> Moreover, when they controlled for other factors commonly associated with first-year grades, the relationship between receiving feedback and grade outcome was larger for the students whose LSAT/UGPA index scores were below the median in their entering law school class than for those with higher predictive scores.<sup>56</sup>

This contrasted with Curcio’s studies, which found a higher correlation between feedback and grades for students with above-median LSAT scores.<sup>57</sup> At least two factors, however, might explain the apparent discrepancy. First, Curcio and her colleagues measured the relationship between feedback and final grades in a single course, while Schwarcz and Farganis assessed the association between feedback in one course and performance in a second course. It is possible that these relationships differ for students with diverse entering credentials.

Second, as Schwarcz and Farganis note, Curcio’s students had lower LSAT scores than the students in their Minnesota study.<sup>58</sup> Indeed, the “above median” students in Curcio’s research had similar LSAT scores to the “below median” students in the

---

<sup>55</sup> *Id.* at 25. From the group that received individualized feedback, 27% scored 0.2 or more below the mean grade in the double section, while 38% of the students from the non-feedback group fell below that benchmark. This difference was statistically significant ( $p = .001$ ). *Id.*

<sup>56</sup> *Id.* at 30, 31 tbl.3. For these analyses, Schwarcz and Farganis used students’ predictive indices, which combine LSAT score and UGPA, as the dependent variable. In a regression equation for students with below-median predictive indices, the coefficient for feedback was .172 ( $p < .05$ ). In the equation for students with above-median predictive indices, that coefficient was just .082 and merely approached statistical significance ( $p < .10$ ).

<sup>57</sup> *Does Practice Make Perfect?*, *supra* note 8, at 294-97; Sargent and Curcio, *supra* note 2, at 391-92.

<sup>58</sup> Schwarcz & Fargani, *supra* note 4, at 30 n.72.

Schwarcz and Farganis study.<sup>59</sup> The results, therefore, could be reconciled to “suggest[] the possibility that feedback might be most beneficial for students within a specific range of LSAT scores.”<sup>60</sup>

In our study, which also capitalizes on a natural experiment, we wanted to test whether individual formative feedback from one professor correlated with student performance in that class as well as in other classes. We also wanted to test which factors, such as LSAT score, UGPA, race, and gender, affected the correlation between receiving feedback and student performance. For this article, our analyses are limited to only one type of feedback, which means that our results will only assess the relationship between that type of feedback and learning outcomes.

### *B. A Natural Experiment at the Moritz College of Law*

The Moritz College of Law divides its first-year J.D. class into three sections. During the period we studied, each section was composed of 54 to 59 first-year students. Students took nine first-year courses; three of them met in smaller subsections.<sup>61</sup> All nine

---

<sup>59</sup> The students at Curcio’s school registered a median LSAT of 159. *Does Practice Make Perfect?*, *supra* note 8, at 293; Sargent & Curcio, *supra* note 2, at 391. The median LSAT score at Minnesota’s law school, in contrast, ranged from 164 to 167 for the classes studied by Schwarcz and Farganis. *University of Minnesota Law School—2015 Standard 509 Information Report*, <http://abarequireddisclosures.org/> (last visited Mar. 10, 2017); *University of Minnesota Law School—2011 Standard 509 Information Report*, <http://abarequireddisclosures.org/> (last visited Mar. 10, 2017).

<sup>60</sup> Schwarcz & Farganis, *supra* note 4, at 30 n.72. Other factors, of course, could also reconcile these results. Those include different types of feedback, different professors, other differences in student achievement or motivation, and the subject matter of the courses studied.

<sup>61</sup> During the fall semester, J.D. students took Legal Analysis and Writing I (LAW I) (2 credits), Torts (4 credits), Criminal Law (4 credits), and Civil Procedure I (4 credits). The legal writing course was taught in subsections of 15–19 students, and each student had one of the other courses taught in a half-section of 27–31 students. In the spring, students took Legal Analysis and Writing II (LAW II) (2 credits), Constitutional Law (4 credits), Contracts (4 credits), Property (4 credits), and Legislation (3 credits). The legal writing course was taught in subsections of 16–19 students, but the other courses all met in their full sections.

The College revised its first-year curriculum in spring 2017, with changes to take effect in spring semester 2018. During the three years we studied, however, the curriculum was as described above.

first-year courses were required; Moritz offered no electives to first-year students at that time.

This scheduling approach, which is common at law schools, facilitates the emergence of natural experiments. Lockstep courses and sections eliminate many of the variables that plague experiments in undergraduate education. Rather than attempt to create control groups in advance, researchers can analyze differences that emerge organically among professors and sections.

This type of natural experiment occurred at Moritz during the last three academic years (2013-14, 2014-15, and 2015-16). Professor Ruth Colker, who taught a section of the first-year Constitutional Law course each spring, offered her students an optional practice exam question. She distributed an essay question from a previous exam and gave students a two-week window to submit an answer.<sup>62</sup> Students who took the practice exam question received prompt feedback; most obtained that feedback within forty-eight hours. The feedback included written comments that were incorporated into the students' practice essays, as well as an estimated numerical grade based on the rubric she used to grade the original exam. The written feedback included comments on how the student approached an exam question as well as specific feedback on how to better use the facts to respond to the particular legal question. Professor Colker also strongly encouraged students to speak with her about the practice exams and offered ample sign-ups for those conferences.

---

<sup>62</sup> Professor Colker's final exam in the course was a take-home consisting of two or three questions. Students each year had 28 hours to complete the final. To model the conditions of the final exam, Professor Colker allowed students to complete the practice essay at home, without time restrictions (other than the two-week window). To simulate the actual exam experience, she encouraged students to spend no more than 8 hours on the practice essay question, which had appeared on a previous year's final exam.

The feedback was purely formative. Students received no course credit for taking the practice exam; nor did the estimated grade contribute to their final grade in the course. Partly because of the formative nature of the exercise, Professor Colker did not grade the practice essays anonymously.

Over the course of the three years, about half of the 167 first-year students enrolled in the class opted to take the practice exam question.<sup>63</sup> This pattern allowed us to compare outcomes for students who took the exam (“takers”) with those who did not (“non-takers”). As we acknowledge below, students chose which of these groups to join; selection bias, therefore, affects some of our analyses. Our ability to control for key student characteristics, however, allows us to explore this bias and shed at least preliminary insight into the relationship between formative feedback and student performance.

We adopted three outcome variables: whether students chose to take the practice exam question; students’ final grade in Professor Colker’s course; and students’ weighted grade-point average in *other* courses taught that spring semester. These three variables allowed us to explore three questions about the feedback offered in this natural experiment:

1. What characteristics predicted a student’s choice to take the practice exam question?
2. Was taking the practice exam question associated with a higher grade in the same course?
3. Was taking the practice exam question associated with higher grades in other courses taught the same semester?

---

<sup>63</sup> Our analyses included only first-year students taking a full course load. A small number of Moritz students “light load” one or more courses during the spring semester of their first year. Five out of the 172 students taking Professor Colker’s course pursued that option; we excluded them from our analysis. Similarly, we excluded any upperlevel or LL.M. students taking Professor Colker’s class.

Moritz grades students on both letter and numerical scales. Letter grades range from E (failing) to A (excellent), with pluses and minuses for some of the letter categories. Numerical grades range from 60 – 100. The scales align as shown in Table 1.

**Table 1: Moritz College of Law Grading Scale**

Letter Grade	Numerical Grade
A	93 – 100
A -	90 – 92
B +	87 – 89
B	83 – 86
B -	80 – 82
C +	77 – 79
C	70 – 76
D	65 – 69
E	60 – 64

We used numerical grades for our outcome measures because they allow finer distinctions. To control for grading differences among professors and over time in classes other than Professor Colker’s Constitutional Law course, we computed Z-scores within each class section.<sup>64</sup> When computing the grade-point average for spring-semester courses other than Constitutional Law, we weighted Z-scores to reflect the credits assigned each course.<sup>65</sup>

---

<sup>64</sup> The Moritz guideline for first-year grades in first-year podium classes at the time of this study was 30% A’s, 60% B’s, and 10% C’s or lower. The curve for LAW I and LAW II courses allowed a higher percentage of As. There was no requirement for the mean grade in a class. Thus, within the guidelines, there is some variation among professors in the distribution of numerical grades within the College’s required curve. Z-scores, calculated as the observed score minus the mean for that class and then divided by the standard deviation for that class, allowed us to standardize distributions and compare mean GPAs across sections and years.

<sup>65</sup> See *supra* note 61.



We included six control variables in our analyses: LSAT score, UGPA, gender, race (white/nonwhite),<sup>66</sup> year of enrollment, and fall-semester GPA. We calculated the latter variable in the same manner that we computed the spring-semester averages: We first calculated Z-scores for students within each class section and then computed a weighted average of those scores. We used SPSS version 23, supplemented by Stata version 14.1, for the analyses reported here.

### III. RESULTS

#### *A. Who Took the Practice Exam Question?*

As Table 2 reflects, students were significantly more likely to take the practice exam question in 2014–15 than in the other two years we studied. Sixty-four percent of students took the exam in 2014–15, while 40–42% of students elected the exam in the other two years. We have been unable to identify a reason for this variation, but we control for it in our multivariable analyses.<sup>67</sup>

**Table 2: Distribution of Students Taking the Practice Exam Question by Year**

<b>Year</b>	<b>Takers</b>	<b>Nontakers</b>	<b>Total Enrolled</b>	<b>Percentage Takers</b>
2013 – 14	23	34	57	40 %
2014 – 15	34	19	53	64 %
2015 – 16	24	33	57	42 %
Total	81	86	167	49 %

When we examined our five other control variables that might contribute to differences in grade outcomes, we found significant differences between takers and non-

---

<sup>66</sup> Although we coded for several racial groups, we did not have enough numbers in any category to allow an analysis more refined than white/non-white.

<sup>67</sup> We treated the first year of results (2013-14) as the reference group for all analyses, and created dummy variables for 2014-15 (“Year Two”) and 2015-16 (“Year Three”).

takers on just two of those measures. As Table 3 reports, takers were significantly more likely to be female; they also had significantly higher UGPA's than non-takers.<sup>68</sup> Notably, the two groups were not significantly different in LSAT score, race, or the grades they earned during the first semester of law school.<sup>69</sup>

**Table 3: Control Variables—Characteristics of Practice Exam Takers and Non-takers**  
(N=167)

Variable	Takers	Non-takers	Sig.
Mean LSAT	159.33	158.80	.452
Mean UGPA	3.66	3.57	<b>.027</b>
Mean Fall-Semester LGPA (weighted Z-scores)	.0743	-.0014	.533
Percent Women	.60	.41	<b>.011</b>
Percent Nonwhite	.19	.21	.696

To analyze more closely the characteristics that distinguished takers from non-takers, we created a logistic regression equation with a binary outcome variable reflecting whether a student took the practice exam question. Table 4 reports the results of these analyses, which were consistent with the bivariate relationships described above.<sup>70</sup>

**Table 4: Logistic Regression for Taking The Practice Exam—Characteristics of Practice Exam Takers**  
(N = 167)

---

<sup>68</sup> For each continuous variable (LSAT score, UGPA, and fall semester LGPA), we examined the difference between takers and non-takers using an independent samples t-test. Levene's test for equality of variances suggested that variances were unequal for our first measure, LSAT scores. Accordingly, we used a t-test adjusted for equal variances not assumed. For the other variables, Levene's test allowed us to assume equal variances. For the two binary variables (women and race), we compared the difference between groups using chi-square test for proportions. For all analyses, we used the conventional .05 level to identify statistically significant results.

<sup>69</sup> The relatively small number of nonwhite students, combined with the racial diversity of that group, counsels caution in interpreting race-related results.

<sup>70</sup> Although LSAT score, UGPA, and fall-semester GPA showed modest correlations, the variance inflation factor was low enough to include all three variables in the same equation (VIF<2 for all). For LSAT and UGPA,  $r = -.158$  ( $p < .05$ ); for LSAT and fall-semester GPA,  $r = .479$  ( $p < .001$ ); and for UGPA and fall-semester GPA,  $r = .303$  ( $p < .001$ ). Together, LSAT score and UGPA predicted about 37% of the variance in fall-semester grades for the students in this population.

	<b>Coefficient</b>	<b>Std. Error</b>	<b>z</b>	<b>Sig.</b>
<b>UGPA</b>	2.026	.796	2.54	<b>0.011</b>
<b>LSAT</b>	.071	.047	1.53	0.126
<b>Fall-Semester GPA</b>	- .251	.294	- 0.85	0.394
<b>Gender</b>	- .777	.337	- 2.30	<b>0.021</b>
<b>Race</b>	.178	.510	0.35	0.727
<b>Year Two</b>	1.170	.421	2.78	<b>0.005</b>
<b>Year Three</b>	.111	.406	0.27	0.786
<b>Constant</b>	- .125	.355	- 0.35	0.726
<b>Pseudo R<sup>2</sup> = 9.63</b>				
<b>Model Significance = .002</b>				

Women were significantly more likely to take the practice exam question, even after controlling for admissions credentials and fall-semester GPA.<sup>71</sup> Fall-semester grades and LSAT did not predict a student's decision to take the exam, but UGPA did; students with a higher UGPA were significantly more likely to take the exam. Race showed no significant relationship to taking the exam.<sup>72</sup> Disproportionate interest in taking the exam during year two, finally, remained relevant even after controlling for other variables.

These results suggest that the students who seek formative feedback differ in some ways from their classmates. In our study, the feedback-seeking students were more likely to be female and to hold higher UGPAs than their classmates. These selection biases affect our analysis of grade outcomes, discussed further below. The findings, however, are equally important in their own right. Why were women more likely than men to take the practice exam question? Why did students with higher UGPAs disproportionately seek this feedback while students with higher LSAT scores did not? Why didn't a student's fall-semester performance show a significant association with

<sup>71</sup> The negative coefficient for gender in both equations reflects the fact that we coded women as "0" and men as "1" on our gender variable.

<sup>72</sup> As explained above, we used a dichotomous variable (white/nonwhite) to designate race/ethnicity. *See supra* note 66. The number of nonwhite students was too small to allow further differentiation.

seeking spring-semester feedback? We discuss those questions in Part IV below.

*B. The Potential for Selection Bias*

As noted above, students chose whether to take the practice exam question. Any significant association between grade outcomes and taking the practice exam question, therefore, might relate to characteristics associated with student choice—rather than the experience of taking the exam itself. Indeed, we identified two significant differences between exam takers and non-takers: the former were more likely to be female and entered law school with higher UGPAs.

It is impossible to eliminate the effects of selection bias, just as it is impossible to eliminate the effects of uncontrolled variables. This fact is part of the reason why statistical analyses of human behavior focus on associations rather than causal relationships and why randomized experiments are necessary to establish causality definitively. We were able, however, to cabin the effects of selection bias in three ways.

First, we controlled for both gender and UGPA in our regression analyses of spring-semester outcomes. We thus controlled for the characteristics that were associated with a student's decision to take the practice exam question. We are able to report whether taking the practice exam question was associated with higher grades, even after controlling for factors associated with a student's decision to take that exam.

Second, we created interaction terms that reflected (a) the interaction between gender and taking the practice exam question, and (b) the interaction between UGPA and taking that exam. These terms allowed us to control for possible differential effects of feedback by gender and UGPA and to look more closely at those relationships.

Finally, we controlled for fall-semester law grades in our analyses of spring-semester outcomes. If higher grades are associated with factors like diligence, good study habits, or a tendency to seek out formative feedback (rather than with formative feedback itself), then those relationships should appear in both the fall and spring semesters of law school. By controlling for fall-semester grades, as well as for an interaction between those grades and a student's choice to take the practice exam question, we attempted to distill the relationships among student characteristics, the experience of taking the practice exam question, and spring-semester grade outcomes.

### *C. Spring Semester Outcomes*

Previous research suggests that formative feedback is associated with improved student performance in both the course offering the feedback<sup>73</sup> and other courses.<sup>74</sup> Our database allowed us to explore both of these relationships: we created one outcome variable reflecting a student's final grade in the Constitutional Law class and a second variable denoting the student's weighted average of grades in other classes taught during the first-year spring semester.

1. *Outcomes in Constitutional Law.* We predicted that students who took the practice exam question in Constitutional Law would earn higher grades on the final exam for that class. Several factors enhanced this possibility: the two tests reflected the same legal field, used a similar format, and were graded by the same person.

---

<sup>73</sup> See *Does Practice Make Perfect?* *supra* note 8, at 299; Negin, *supra* note 9, at 676; Sargent & Curcio, *supra* note 2, at 391-92.

<sup>74</sup> See Schwarcz & Farganis, *supra* note 4. The association between feedback and exam performance carried over from fall to spring semester. In the spring-semester double sections, students who had received feedback outperformed their peers who had not received it even when the feedback occurred during a fall-semester class. *Id.* at 20 tbl.1, 23 fig.3. But see *Does Practice Make Perfect?* *supra* note 8, at 306-07 (finding no association between feedback in Civil Procedure and grades in students' other spring semester courses).

Table 5 reveals that students who took the practice exam did, on average, earn higher grades on the final exam in that class. In this unadjusted analysis, the difference between the two groups was 2.6 points on Moritz's numerical grading scale ( $p = .005$ ). At the mean, that gap translated to an average grade of B-plus for students who took the practice exam, and B for those who did not.<sup>75</sup>

**Table 5: Grades on Final Exam,  
Practice Exam Takers v. Nontakers**

Practice Exam Taker	Number of Students	Mean	Std. Deviation	Std. Error Mean
No	86	85.22	5.828	.628
Yes	81	87.85	6.191	.688

As discussed above, takers and non-takers did not differ significantly on their LSAT scores, first-semester law school grade-point average or racial composition.<sup>76</sup> They did, however, differ in their gender mix, UGPAs, and year of enrollment in the class. To control for these differences, we estimated a regression equation using the students' final exam grade as a dependent variable. For independent variables, we included LSAT score, UGPA, fall-semester law school GPA, gender, race, the year that a student enrolled in the class,<sup>77</sup> and whether that student took the practice exam question.<sup>78</sup>

Table 6 reports the results of this regression analysis. The coefficient for fall-semester GPA shows a strong association with grades on Professor Colker's final exam.

---

<sup>75</sup> See Table 1 *supra*. We do not present Z-scores for our analysis of final grades in Professor Colker's course in Table 5 because she graded all of the exams in that analysis. Moreover, using Z-scores revealed the same statistically significant difference. Students who took the practice exam averaged z scores of .239, while those who did not take the practice test averaged -.204 ( $p = .004$ ).

<sup>76</sup> See *supra* note 69 and accompanying text.

<sup>77</sup> See *supra* note 67 for a description of how we constructed dummy variables to control for year.

<sup>78</sup> Although LSAT score, UGPA, and fall-semester GPA showed modest correlations, they did not demonstrate unacceptable levels of collinearity; thus, we were able to include them in the same equation. See also *supra* note 70.

The only other coefficient that reaches significance is the one for taking the practice exam question: students who took that exam received significantly higher grades on the course final, even after controlling for their prior academic credentials. The grade bump associated with taking the practice exam was slightly smaller in the regression equation (2.3 points) than in our unadjusted analysis, but it was still practically significant. At many points on the Moritz grading scale, a 2.3 point difference would equal a half-grade change in letter grade.

**Table 6: Linear Regression for Grade on Professor Colker's Final Exam**  
(N = 167)

	<b>Coefficient</b>	<b>Std. Error</b>	<b>t</b>	<b>Sig.</b>
<b>Taker</b>	2.316	.851	2.72	<b>0.007</b>
<b>UGPA</b>	2.154	1.857	1.16	0.248
<b>LSAT</b>	.149	.111	1.35	0.180
<b>Fall-Semester GPA</b>	3.802	.708	5.37	<b>0.000</b>
<b>Gender</b>	.545	.825	0.66	0.510
<b>Race</b>	1.281	1.219	1.05	0.295
<b>Year Two</b>	.147	1.017	0.14	0.885
<b>Year Three</b>	1.527	.982	1.55	0.122
<b>Constant</b>	84.282	.956	88.14	<b>0.000</b>
<b><i>Adjusted R<sup>2</sup> = .298</i></b>				
<b><i>F = 9.79</i></b>				
<b>Significance of <i>F</i> &lt; .0001</b>				

The significance of the practice exam coefficient is particularly notable, because the equation controls for fall-semester law school grades. If the positive association between the practice exam and final grade stemmed from selection bias, we would expect including fall-semester grades in the equation to reduce that association. Characteristics like diligence and good study habits, which might prompt a student both to take an optional practice exam and to obtain a higher final course grade, would have also influenced fall-semester grades. Including fall-semester grades in the equation, therefore,

should control for some aspects of selection bias. The coefficient for taking the practice exam, however, remained significant when we controlled for fall-semester grades.

Adding an interaction term (fall-semester grades \* taker) to the equation did not change these results. The coefficients for (a) fall-semester grades and (b) taking the practice exam question remained significant and positive, while the interaction term lacked significance ( $p = .248$ ).<sup>79</sup> The average benefit from taking the practice exam thus was similar regardless of fall-semester grades; students with low and high fall-semester grades who took the practice exam experienced a similar average increase in the final course grade.

Similarly, introduction of an interaction term for gender and taking the practice exam question did not affect the regression results. The coefficient for taking the practice exam question remained significant; the coefficient for gender remained nonsignificant; and the interaction term was not significant ( $p = .226$ ).<sup>80</sup> These results suggest that, although women were more likely than men to take the practice exam, both women and men (on average) secured higher final grades in the class if they took advantage of the practice exam opportunity. The significant relationship between taking the practice exam question and final grade was not explained by a selection bias based on gender.

Given the relationship between UGPA and taking the practice exam (Table 4), we also tested an interaction term for UGPA and taking that exam. Once again, our regression results remained stable. The coefficient for taking the practice exam remained significant; the one for UGPA remained nonsignificant; and the coefficient for the interaction term lacked significance ( $p = .626$ ).

---

<sup>79</sup> Full results of these analyses are available from the authors upon request.

<sup>80</sup> Full results of these analyses are available from the authors upon request.



We then used our database to approximate the median-split analyses performed by Curcio and her colleagues.<sup>81</sup> When we divided students by median UGPA, we found that the coefficient for taking the practice exam was larger for students with below-median UGPAs than for those with a UGPA at or above the median. Indeed, the coefficient for the latter group did not reach conventional significance levels.<sup>82</sup> Those results, depicted in Tables 7 and 8, differ from the ones reported by Curcio and her colleagues: those researchers found a significant association between feedback and final exam grades only for students with *higher* UGPAs.<sup>83</sup>

**Table 7: Linear Regression for Grade on Professor Colker's Final Exam  
Limited to Students with Below-Median UGPAs  
(N = 80)**

	<b>Coefficient</b>	<b>Std. Error</b>	<b>t</b>	<b>Sig.</b>
<b>Taker</b>	3.410	1.051	3.24	<b>0.002</b>
<b>LSAT</b>	.5382	.148	3.64	<b>0.001</b>
<b>Fall-Semester GPA</b>	2.330	.897	2.60	<b>0.011</b>
<b>Gender</b>	.1792	1.044	0.17	0.864
<b>Race</b>	1.184	1.361	0.87	0.387
<b>Year Two</b>	1.599	1.271	1.26	0.213
<b>Year Three</b>	2.850	1.234	2.31	<b>0.024</b>
<b>Constant</b>	82.123	1.308	62.79	<b>0.000</b>
<b><i>Adjusted R<sup>2</sup> = .372</i></b>				
<b><i>F = 7.70</i></b>				

<sup>81</sup> See *supra* notes 31 & 43 and accompanying text. We also attempted to replicate the analyses performed by Schwarcz and Farganis, who used the median predictive index (a combination of LSAT and UGPA) to divide their subjects into two groups. We applied that analysis, as Schwarcz and Farganis did, to examine the association between feedback and performance in other courses. See *infra* notes 93–95 and accompanying text.

<sup>82</sup> At first glance, this result seems in tension with the finding reported in the previous paragraph: that a coefficient reflecting the interaction between UGPA and taking the practice exam failed to attain significance. There are, however, two explanations for this. First, the positive relationship between taking the practice exam and scoring a higher grade in the class might have been focused in the bottom half of the class (by UGPA) without being linear. Second, the smaller sample sizes analyzed in these median-split equations makes it harder for coefficients to attain statistical significance.

<sup>83</sup> Curcio's students had a median UGPA of 3.4, see *Practice Makes Perfect*, *supra* note 8, at 293, while our three classes registered a median UGPA of 3.60 to 3.65. Curcio's study, however, occurred almost a decade before our analysis. Given the prevalence of college grade inflation, we do not place much weight on the difference in median UGPAs.

Significance of  $F < .0001$

**Table 8: Linear Regression for Grade on Professor Colker's Final Exam  
Limited to Students with UGPAs at or Above the Median  
(N = 87)**

	<b>Coefficient</b>	<b>Std. Error</b>	<b>t</b>	<b>Sig.</b>
<b>Taker</b>	2.257	1.335	1.69	0.095
<b>LSAT</b>	-.156	.155	-1.01	0.316
<b>Fall-Semester GPA</b>	5.160	.995	5.19	<b>0.000</b>
<b>Gender</b>	.2859	1.253	0.23	0.820
<b>Race</b>	1.162	2.180	0.53	0.596
<b>Year Two</b>	-.928	1.558	-0.60	0.553
<b>Year Three</b>	-.0195	1.484	-0.01	0.990
<b>Constant</b>	85.280	1.375	62.01	<b>0.000</b>
<i>Adjusted R<sup>2</sup> = .265</i>				
<i>F = 5.42</i>				
<b>Significance of <math>F &lt; .0001</math></b>				

When we divided students by median LSAT, our results also differed from those of Curcio and her colleagues. In our database, the coefficient for taking the practice exam was significant for *both* students with below-median LSAT scores and those with scores at or above the median (Tables 9 & 10). As reported above, Curcio and her colleagues found a significant association between feedback and final course grade only for students with LSAT scores at or above the median.<sup>84</sup> This difference is intriguing because the median LSAT for our students (159) was identical to the median LSAT among Curcio's students.<sup>85</sup>

<sup>84</sup> See *supra* note 31 & 43 and accompanying text.

<sup>85</sup> See *supra* note 59. Table 10 reveals another intriguing relationship between race and final course grade. We discuss that relationship *infra* notes 100-101 and accompanying text.

**Table 9: Linear Regression for Grade on Professor Colker's Final Exam  
Limited to Students with Below-Median LSAT Scores  
(N = 77)**

	<b>Coefficient</b>	<b>Std. Error</b>	<b>t</b>	<b>Sig.</b>
<b>Taker</b>	3.116	1.485	2.10	<b>0.040</b>
<b>UGPA</b>	.333	3.419	0.10	0.923
<b>Fall-Semester GPA</b>	4.978	1.183	4.21	<b>0.000</b>
<b>Gender</b>	.170	1.362	0.12	0.901
<b>Race</b>	-.198	1.763	-.011	0.911
<b>Year Two</b>	-.060	1.702	-.004	0.972
<b>Year Three</b>	.775	1.607	0.48	0.631
<b>Constant</b>	84.809	1.518	55.87	<b>0.000</b>
<i>Adjusted R<sup>2</sup> = .253</i>				
<i>F = 4.68</i>				
<i>Significance of F = .0002</i>				

**Table 10: Linear Regression for Grade on Professor Colker's Final Exam  
Limited to Students with LSAT Scores at or Above the Median  
(N = 90)**

	<b>Coefficient</b>	<b>Std. Error</b>	<b>t</b>	<b>Sig.</b>
<b>Taker</b>	2.225	1.046	2.13	<b>0.037</b>
<b>UGPA</b>	3.137	2.199	1.43	0.157
<b>Fall-Semester GPA</b>	3.572	.893	4.00	<b>0.000</b>
<b>Gender</b>	1.596	1.077	1.48	0.142
<b>Race</b>	3.530	1.737	2.03	<b>0.045</b>
<b>Year Two</b>	.144	1.237	0.12	0.908
<b>Year Three</b>	2.106	1.246	1.69	0.095
<b>Constant</b>	83.904	1.285	65.29	<b>0.000</b>
<i>Adjusted R<sup>2</sup> = .250</i>				
<i>F = 5.23</i>				
<i>Significance of F = .0001</i>				

Taking Professor Colker's practice exam, in sum, was associated with earning a higher grade on her final exam. The increase was both statistically and practically significant, spelling the difference between an average grade of B and one of B-plus. The relationship persisted after controlling for students' entering credentials and first-

semester law grades. The positive association was stronger for students with below-median UGPAs than for those with higher UGPAs, but it existed for students with LSAT scores both below and at/above the class median.

2. *Outcomes in Other Classes.* Some studies suggest that formative feedback is associated with student performance in classes other than the one in which feedback was obtained. Researchers at the University of Minnesota Law School, for example, found that feedback in one first-year course was associated with higher grades in other courses.<sup>86</sup> We were able to explore this effect in our database by comparing the grades that takers and non-takers earned in spring-semester courses *other than* the class in which they took the practice exam question.

The students in our database took four other courses during the semester they enjoyed the option of taking the practice exam question. As Table 11 reflects, students who completed the practice exam averaged higher grades in all four of these courses;<sup>87</sup> in two of the courses, the difference was statistically significant.<sup>88</sup> When we created a weighted average of these four grades, we also found a significant difference between takers and non-takers; figures for the weighted average appear in the bottom line of Table 11.<sup>89</sup>

---

<sup>86</sup> Schwarcz & Farganis, *supra* note 4, at 29-30, 31 tbl.3.

<sup>87</sup> To allow comparisons between professors and over time, grades are expressed as z-scores, a standardized measure of the distance from the mean grade that takes account of the standard deviation.

<sup>88</sup> In one course, LAW II, students completed a small portion of their graded work before taking the practice exam in Professor Colker's course. The former work, however, comprised no more than 10% of the student's grade so we did not distinguish between LAW II and other spring-semester courses.

<sup>89</sup> The weighted average reflected the credits assigned each course. *See supra* note 61.

**Table 11: Spring-Semester Grades for Takers  
And Non-Takers**

<b>Course</b>		<b>Number of Students</b>	<b>Mean Z-Score</b>	<b>Std. Deviation</b>	<b>Std. Error Mean</b>	<b>Sig.</b>
<b>Contracts</b>	Non-taker	86	-.137	.853	.092	<b>.023</b>
	Taker	81	.209	1.088	.121	
<b>Property</b>	Non-taker	86	-.052	1.020	.110	.343
	Taker	81	.095	.980	.109	
<b>Legislation</b>	Non-taker	86	-.041	.957	.103	.586
	Taker	81	.043	1.036	.115	
<b>LAW II</b>	Non-taker	86	-.133	1.001	.108	<b>.024</b>
	Taker	81	.207	.926	.103	
<b>Weighted Average</b>	Non-taker	86	-.113	.748	.081	<b>.033</b>
	Taker	81	.153	.856	.095	

We further explored the relationship between taking the practice exam and grades in other courses by creating a regression equation similar to the one reported in Table 6. For this analysis we used the weighted average of grades earned in spring-semester courses other than Constitutional Law as our dependent variable; the independent variables mirrored those in Table 6. Table 12 reports the result of this regression analysis.

**Table 12: Linear Regression for Weighted Spring-Semester GPA (In Courses Other Than Constitutional Law)**  
(N = 167)

	<b>Coefficient</b>	<b>Std. Error</b>	<b>t</b>	<b>Sig.</b>
<b>Taker</b>	.175	.079	2.22	<b>0.028</b>
<b>UGPA</b>	.234	.172	1.36	0.176
<b>LSAT</b>	.021	.010	2.04	<b>0.043</b>
<b>Fall-Semester GPA</b>	.747	.066	11.37	<b>0.000</b>
<b>Gender</b>	.159	.076	2.08	<b>0.039</b>
<b>Race</b>	.065	.113	0.57	0.566
<b>Year Two</b>	- .023	.094	- 0.24	0.812
<b>Year Three</b>	.038	.091	0.42	0.676
<b>Constant</b>	- .163	.089	- 1.84	<b>0.068</b>
<i>Adjusted R<sup>2</sup> = .640</i>				
<i>F = 37.90</i>				
<b>Significance of F &lt; 0.0001</b>				

The variables in this regression analysis explain an impressive amount of the variance (64.0%) in spring-semester grades other than Constitutional Law. This degree of explanatory power is unusual in social science research. Fall-semester law school grades accounted for the lion’s share of that variance, but the coefficients for LSAT score, taking the practice exam question, and gender were also significant. In our study, as in the one by Schwarcz and Farganis, feedback in one course showed a positive relationship with grades in other courses—even after controlling for other factors.

The size of this relationship, moreover, was practically significant. Students who completed Professor Colker’s practice exam achieved a weighted GPA in their other spring-semester classes that averaged .83 points (on the Moritz scale) more than the grades of their classmates.<sup>90</sup> Grade compression among law students means that a difference of this size noticeably affects GPA and class rank. The students in our

<sup>90</sup> For the Moritz grading scale, see *supra* Table 1.

database, for example, had a median GPA of 87.74 at the end of their first year. A bump of .83 points would have increased the median student's GPA to 88.57, moving her to the fifty-ninth percentile.<sup>91</sup> The same bump would have moved a student from the thirtieth percentile to the forty-first, or from the seventy-fifth percentile to the eighty-first.<sup>92</sup>

Schwarcz and Farganis found that the relationship between formative feedback and grades in other courses was markedly stronger for students with predictive indices below their class median than for classmates with indices at or above the median.<sup>93</sup> We uncovered a similar difference: When we divided our students by median predictive index and analyzed grades in spring-semester courses other than Constitutional Law, the coefficient for taking the practice exam was somewhat larger for students with below-median indices (2.68) than for those with indices at or above the median (2.30).<sup>94</sup> The difference, however, was much more modest than the one noted by Schwarcz and Farganis. The smaller size of our population, moreover, counsels caution in placing too much weight on these results; the coefficient for taking the practice exam merely

---

<sup>91</sup> These illustrations rely on calculations using only the students included in the regression analysis reported in Table 12. In unreported analyses we confirmed that the magnitude of the effect is similar when we calculate GPA as the registrar's office does, for all students enrolled in the first-year class

<sup>92</sup> Moritz, like some other law schools, releases only limited information about class rank. The College computes individual ranks for students in the top 5% of each class and then publishes approximate grade cut-offs for selected brackets of the class. *See* Moritz College of Law, Faculty Rules of the College of Law, Rule 9.11 (July 1, 2016), <http://moritzlaw.osu.edu/registrar/wp-content/uploads/sites/15/2016/08/Moritz-College-of-Law-Faculty-Rules-20161.pdf>. A student, therefore, could not report that her GPA placed her precisely at the 41<sup>st</sup>, 59<sup>th</sup>, or 81<sup>st</sup> percentile. Students and employers, however, can read between the published cut-off lines to approximate a student's rank; indeed, an increase of .83 points would move some students from one bracket to a higher one. An increase of this magnitude almost certainly would affect assessment of a student's rank.

<sup>93</sup> *See supra* note 56 and accompanying text. For the first group, Schwarcz and Farganis reported a coefficient of .172 ( $p < .05$ ); for the second, it was .082 ( $p < .10$ )

<sup>94</sup> For the first coefficient,  $p = .092$ ; for the second,  $p = .036$ . In both of these equations, we used the independent and dependent variables reported in Table 12. Full results are available from the authors upon request.

approached significance in the equation for students with below-median predictive indices.<sup>95</sup>

### *C. Gender*

Our analyses revealed two intriguing gender differences. Women were significantly more likely than men to take the optional practice exam question, even after controlling for UGPA, LSAT score, and fall-semester law grades.<sup>96</sup> Men, conversely, appeared to receive higher grades in some spring-semester courses. As Table 12 reveals, the association between gender and weighted spring-semester average was significant after controlling for the other factors in that equation.<sup>97</sup>

We further explored the latter effect by examining each spring-semester course separately. We already knew that gender was not significantly associated with final grades in Professor Colker's course (Table 6). By applying the equation from that table to standardized final grades in each other spring-semester course we determined that gender was significantly associated with final grades in Property ( $p = .002$ ), Legislation ( $p = .008$ ), and LAW II ( $p = .005$ ).<sup>98</sup> Final grades in the fifth spring-semester course, Contracts, showed no significant relationship to gender ( $p = .496$ ) in the regression analysis.

---

<sup>95</sup> See *supra* note 94. Schwarcz and Farganis analyzed data for 558 students, with 278 falling in their "below median" group and 280 in the "above median" category. Schwarcz & Farganis, *supra* note 4, at 31 tbl.3. In contrast, we had only 167 students in our full analyses; when we divided them by median predictive index, the subgroups included just 59 below-median students and 108 at or above the median.

<sup>96</sup> See *supra* notes 68-71 and accompanying text.

<sup>97</sup> As noted above, *supra* note 71, we coded our gender variable "0" for women and "1" for men. The positive coefficient in Table 12, therefore, suggests that men obtained higher spring-semester averages than women after controlling for other factors.

<sup>98</sup> We used standardized grades (z-scores) for these analyses because the professors for these courses varied over the three years we studied. The independent variables in each of these regression equations were the same as those for the equation reported in Table 6: LSAT score, UGPA, fall-semester law school GPA, whether the student took the practice exam, race, gender, and year of enrollment.



The gender differences in Property, Legislation, and LAWII pointed in different directions. Men obtained significantly higher grades in the first two courses, while women secured significantly higher grades in LAW II. The effects remained even after we eliminated the variable for taking the optional practice exam question, so the differences do not seem to stem from women's proclivity to take that exam. Instead, final grades for these students were highly gendered in three of five spring-semester classes.<sup>99</sup>

#### *D. Race*

In our regression analysis, race was not significant in predicting whether a student would take the optional practice exam question. Nor was it significant in our primary analysis of final grades in Professor Colker's course (Table 6). When we limited the latter analysis to students with LSAT scores at or above the median, however, the coefficient for race was statistically significant (Table 10). Nonwhite students in that group achieved significantly higher grades than their white classmates—once we controlled for the other variables noted in Table 10.

The small number of nonwhite students in our population counsels caution in interpreting this result.<sup>100</sup> The outcome is encouraging, however, given the number of studies reporting that nonwhite law students obtain lower grades than their white classmates after controlling for LSAT score, UGPA, and other factors.<sup>101</sup> The results

---

<sup>99</sup> When we expanded this analysis to a larger population, including all first-year students enrolled at Moritz during the last three years (rather than just those in Professor Colker's section), the relationship between gender and final grade remained in the Legislation course and LAW II. It disappeared, however, in the Property course.

<sup>100</sup> Thirteen students (13% of the total in this analysis) were nonwhite. The group included Asian-American, Latino/a, Black, and "other race" students.

<sup>101</sup> See, e.g., John Fordyce et al., *Predicting First-Year Law School Performance: The Influences of Race, Gender, and Undergraduate Major*, 43 EASTERN ECON. J. 64 (2017); Alexia Brunet Marks & Scott A. Moss, *What Predicts Law Student Success? A Longitudinal Study Correlating Law Student Applicant Data*

from Professor Colker's Constitutional Law course suggest that minority students can match—and even exceed—the academic achievement of their white classmates.

#### IV. DISCUSSION

Our case study offers insights on five different questions: (1) Do students who pursue optional feedback differ from those who do not? (2) When students seek and obtain formative feedback in a course, does that factor correlate with a higher final grade in the course? (3) Does the experience of seeking and obtaining feedback correlate with a higher final grade in other courses taught that semester? (4) How does gender affect the answers to these questions? (5) How does race affect the answers?

We explore these questions in the subsections below. We also identify questions our study poses for future research.

##### *A. Who Chooses Formative Feedback?*

The students in our natural experiment decided whether to take the optional practice exam question. Students who exercised that option invested significant time in the experience. Students typically spent time reviewing their casebook and notes before taking the practice exam. They then devoted, on average, about eight hours to writing a practice exam answer. After receiving written feedback, they also presumably spent time reviewing that feedback. Finally, Professor Colker strongly encouraged students to meet with her individually about their answer, and nearly every student did so.

---

*and Law School Outcomes*, 13 J. EMPIRICAL LEGAL STUDS. 205 (2016); Schwarcz & Farganis, *supra* note 4, at 28.

Given this time commitment, it is interesting to consider what type of student chose to take the practice exam. Our finding that female students disproportionately took the practice exam is consistent with a prior study of students in a microeconomics class at Stockholm University.<sup>102</sup> There, female students were more likely to take optional quizzes and do the extra work entailed in earning bonus points for attending seminars.<sup>103</sup> Quiz taking was associated with higher final exam scores.<sup>104</sup> Laboratory experimental work suggests that female students may be more risk averse, more self-disciplined in their study habits, and less overconfident than males.<sup>105</sup> The authors of the Swedish study argued that their classroom results supported higher levels of risk aversion by female students, but not less overconfidence than males. The authors could not rule out an effect of self-discipline based on greater procrastination by males.<sup>106</sup>

We can only speculate on the mechanisms supporting our gender results. Higher levels of risk-aversion and more self-discipline may explain female students being more likely to take the practice exam. The effect of potential overconfidence, however, is a more complicated explanatory factor when it comes to law students. Most empirical educational research supports a link between high-achievement and perceptions of confidence and control.<sup>107</sup> From this perspective, if confidence based on strong fall-semester performance were an important part of the explanation for not taking the practice test, we would expect a positive relationship between higher fall-semester grades

---

<sup>102</sup> Lena Nekby et al., *Gender Differences in Examination Behavior*, 53 ECON. INQUIRY 352 (2015).

<sup>103</sup> *Id.* at 353, 358 tbl.3. One difference from our study was that the scores on the quizzes were credited to the final exam in the course. But taking the quizzes was low risk in that students could choose to redo the quiz questions during the final exam. *Id.* at 354-55.

<sup>104</sup> *Id.* at 359.

<sup>105</sup> *Id.* at 352 (citing studies).

<sup>106</sup> *Id.* at 360-61.

<sup>107</sup> Leah M. Christensen, *Enhancing Law School Success: A Study of Goal Orientations, Academic Achievement and Declining Self-Efficacy of Our Law Students*, 33 LAW & PSYCHOL. REV. 57, 71-73 (2009).

and non-takers of the practice exam. However, as shown in Table 3, students who took the practice exam actually had higher grades in the fall on average than non-takers and, in any event, the correlation between fall grades and taking the practice test was not significant. But one study of law students suggests that we should not assume a connection between strong performance and subsequent high levels of student confidence. In that study, higher-ranking students were *less* likely to believe that they could do the work and meet the academic challenges of law school.<sup>108</sup> If this result from one law school is generalizable, it could suggest a link between strong fall-semester performance and less overconfidence, which would be consistent with a willingness to undertake the work involved in the voluntary practice exam.

Our UGPA findings are also arguably consistent with the hypothesis that choosing to take the practice exam question is associated with better study habits (which may in turn be associated with self-discipline). We found that students who took the voluntary practice exam question were likely to have higher UGPAs but not higher LSAT scores.

Another hypothesis is that students who took the practice exam might place a higher value on formative feedback due to a growth mindset. Carol Dweck's research on mindsets theorizes two categories: growth and fixed.<sup>109</sup> A person with a growth mindset believes that intelligence and skills can be acquired through effort and training.<sup>110</sup> In

---

<sup>108</sup> *Id.* This finding was based on correlations between survey responses of law students and their academic performance. *Id.* Higher performing students also were more likely to be mastery-oriented learners than performance-oriented. *Id.* at 67-68, 75-76. Professor Christensen suggested that their lack of self-confidence may be related to the predominant goal orientation in legal education, which she argues is more focused toward performance than mastery. *Id.* at 78-80.

<sup>109</sup> CAROL S. DWECK, MINDSET 6-7 (2006).

<sup>110</sup> *Id.* at 7.

contrast, a person with a fixed mindset sees these attributes as given and fixed.<sup>111</sup> Having a fixed mindset places an emphasis on performance to demonstrate one's abilities, which means that students with a fixed mindset "are less likely to seek out new learning for fear that it might reveal a weakness."<sup>112</sup> This attitude might have been particularly influential in deterring students from participating if they also felt unprepared to take the mid-semester practice test.

In addition, mindset may have affected how students viewed the utility of the practice exam. Those with a growth mindset are more likely to see a self-test as an opportunity for learning.<sup>113</sup> Students with a fixed mindset, however, are more likely to see such testing as merely a way to check knowledge.<sup>114</sup> Perhaps students with a fixed mindset would be less motivated to do the work entailed in taking a practice test for the purpose of demonstrating their knowledge than students with a growth mindset who would be more likely to see the exam as a step in their learning process. Given our findings on gender and UGPA, this hypothesis prompts the question whether growth mindset might be correlated with higher UGPA? Or with being female?

Finally, one has to wonder if the gender association relates to the fact that Professor Colker is a female. Few studies examine the interactions among student gender, professor gender, and student achievement in higher education.<sup>115</sup> At least one, however,

---

<sup>111</sup> *Id.* at 6.

<sup>112</sup> Elizabeth Ruiz Frost, *Feedback Distortion: The Shortcomings of Model Answers as Formative Feedback*, 65 J. LEGAL EDUC. 938, 953 (2016) (citing Carol S. Dweck & Allison Master, *Self-Theories Motivate Self-Regulated Learning*, in MOTIVATION AND SELF-REGULATED LEARNING 31, 48 (Dale H. Schunk & Barry J. Zimmerman, eds. 2008)).

<sup>113</sup> Veronica X. Yan et al., *Habits and Beliefs That Guide Self-Regulated Learning: Do They Vary With Mindset?*, 3 J. APPLIED RES. IN MEMORY & COGNITION 140, 146 (2014).

<sup>114</sup> *Id.*

<sup>115</sup> See Susan A. Basow et al., *The Effects of Professors' Race and Gender on Student Evaluations and Performance*, 47 COLLEGE STUDENT J. 352, 354-55 (2013) ("The effects of professor race and gender on student learning have rarely been examined.").

found that graduate students feel greater psychosocial comfort with professors of the same gender.<sup>116</sup> It is possible that Professor Colker's female students felt more comfortable interacting with her and sought to extend those opportunities by taking the practice exam question.<sup>117</sup> Alternatively, Professor Colker may have served as a role-model in a way that encouraged female students disproportionately to take the practice exam question.

#### *B. Formative Feedback and Course Grade*

This study reflects a limited examination of three years of experience with one type of voluntary feedback opportunity given to students in one podium class during the spring semester where the final exam was a 28-hour take-home exam and the class was taught by a female professor. It suggests that taking advantage of the formative assessment opportunity was associated with higher grades in that class as well as other spring semester classes. This finding, especially when combined with previous work on the positive effects of formative feedback, supports attempts to increase the availability of that feedback in legal education. The study, however, also raises many questions. Would the results be similar with different kinds of formative assessments under different conditions? We outline below some factors to consider in future research.

In context, this formative feedback opportunity was one instance of feedback in a series received by first-year law students. At Moritz, all law students are assigned to a "small section" of a podium class during the fall semester in which they take a mid-term

---

<sup>116</sup> Debra S. Schroeder & Clifford R. Mynatt, *Graduate Students' Relationships with Their Male and Female Major Professors*, 40 SEX ROLES 393 (1999). See also *infra* notes 157-166 and accompanying text.

<sup>117</sup> Professor Colker has lunch with students on a voluntary basis and finds that female students disproportionately sign up to have lunch with her.

exam for a small part of their grade in that course. All students receive feedback in the form of that mid-term grade; many professors offer additional feedback in the form of classroom review of the exam, model answers, comments on exam papers, and/or individual meetings to discuss the exam. In some small and large sections, students prepare other assignments with written or oral feedback. Students also receive extensive written and oral feedback on legal memoranda in their LAW I class.<sup>118</sup> At the start of the spring-semester, students receive feedback in the form of their fall-semester grades. Some also have the opportunity to review written comments on their exam or to meet individually with the professor to discuss their performance.<sup>119</sup>

The literature supports a hypothesis that repetitive formative feedback may have a cumulative effect on student performance. In the law school setting, Negin's 1981 study found a positive relationship between the number of tests given during the semester and performance on the final.<sup>120</sup> In the Minnesota study, double sections were not included in the primary analysis if students in both component sections received individualized feedback. Of interest here is the finding from one of these double sections that the students who had received prior feedback in two separate classes outperformed those who received feedback in only one class.<sup>121</sup> Thus, in both studies, a "second dose" of feedback

---

<sup>118</sup> The Minnesota study considered only feedback in doctrinal courses and excluded assignments that required students to produce "practical" documents such as complaints or contracts on the theory that these exercises "develop skills that are distinct from those that are tested on traditional law school exams." Schwarcz & Farganis, *supra* note 4, at 16-17. In contrast, we view the analytical work in the LAW I and II classes as highly relevant for developing the skills necessary for a strong performance on final examinations.

<sup>119</sup> A professor can use the same assessment for both formative and summative feedback. Frost, *supra* note 112, at 943. While final examinations provide summative assessment, reviewing them with students can provide lessons that may serve as formative assessment in subsequent classes.

<sup>120</sup> Negin, *supra* note 9, at 675-76; *see supra* text accompanying notes 11-20.

<sup>121</sup> Schwarcz & Farganis, *supra* note 4, at 26-27, 32 n.73. The students in the component section with two feedback experiences had a mean GPA in the double section course of 3.383 compared to 3.284 for those who received feedback in only one of their classes. *Id.*

was associated with better performance. These outcomes are consistent with educational theory that emphasizes the benefits of spaced repetition of key ideas as a way to maximize the effects of retrieving concepts from memory.<sup>122</sup>

The significant effect of Professor Colker's formative assessment on the grades students achieved in her class suggests that, even after multiple other forms of feedback, adding another feedback opportunity provided positive gains. Although the natural experiment reported here does not permit us to compare different types of formative assessment, it is worth considering what attributes may have made this voluntary practice exam exercise so effective. Her practice exam used the same take-home format the students faced on her final, the exercise was voluntary and ungraded, her feedback was prompt, and she used individualized written and verbal comments.

1. *Format of the exercise: take-home practice test.* Professor Colker's summative assessment in Constitutional Law was a 28-hour take-home examination. Her formative assessment exercise matched this format.<sup>123</sup> While some of the other professors who taught spring-semester courses at Moritz used take-home examinations during the years studied, only one of them taught sections in common with Professor Colker, and no

---

<sup>122</sup> See PETER C. BROWN ET AL., MAKE IT STICK: THE SCIENCE OF SUCCESSFUL LEARNING 3-4 (2014). Retrieval that is spaced over time produces longer-lasting learning and promotes application of the learning in new settings. *Id.* at 4.

<sup>123</sup> The Negin study also matched the format of the feedback exercises to that of the final, summative assessment. All were multiple-choice examinations. Negin, *supra* note 9, at 674-75. For the first study by Professor Curcio, both the formative assessments and the final exam were essay questions, although the formative assessment answers were written at home and the final was a traditional timed in-class exam. *Does Practice Make Perfect?*, *supra* note 8, at 287-89. The questions on the final exam were designed to evaluate skills that had been emphasized in the practice writing exercises: students' ability "to break a rule into its component parts, to recognize relevant facts, and to analyze and apply the facts to the applicable constituent element." *Id.* at 289. In Curcio's second study, the formative assessment vehicles varied. They included in-class timed quizzes and take-home quizzes with short-answer and essay questions. There was a partial match in that short-answer questions on the final closed-book timed exam were used to compare the performance of the classes. Sargent & Curcio, *supra* note 2, at 386-89. For the Minnesota study, the formative assessments took place in many different classes and varied greatly, *see supra* text accompanying notes 44-47, and no information is reported on the format of the summative assessments in the double sections.



professors in fall-semester classes used this format. Thus, the practice exam would have been the students' first experience with a take-home format prior to taking Professor Colker's final examination. Students who took the practice exam, moreover, received feedback about both their own performance and about Professor Colker's expectations; they would have been able to use both insights to tailor their preparation for the final exam.

Part of the positive association between taking the practice question and performance on the final exam, therefore, may have been due to exposure to this exam format. This is consistent with empirical studies in other disciplines finding that practice tests were most strongly associated with improved performance if they closely matched the format and difficulty of the actual exams.<sup>124</sup> It also means, however, that the study can be criticized for including an effect unrelated to promoting learning of the substantive course material.<sup>125</sup> From a pedagogical perspective, however, familiarity with the exam format could be a positive effect of taking a practice test. It could mean that the final examination is more likely to test the underlying material and its application, rather than yielding outcomes that rest on a student's facility with the particular exam format. From this perspective, if the exercise fulfilled its goals, students should have gained both a greater understanding of the underlying material in the Constitutional Law course *and* improved test-taking skills.

---

<sup>124</sup> See, e.g., Rene Oliver & Robert L. Williams, *Direct and Indirect Effects of Completion Versus Accuracy Contingencies on Practice-Exam and Actual-Exam Performance*, 14 J. BEHAV. EDUC. 141, 142 (2005) (citing studies showing that taking practice exams similar in format and content to an exam are associated with stronger performance on such exams).

<sup>125</sup> See Schwarcz & Farganis, *supra* note 4, at 13 ("[I]t is hard to know from [Negin's] study design whether the students who performed better after taking multiple exams did so merely because they became familiar with the question types favored by the instructor, rather than learning to understand the underlying material better.")

In fact, one of the rationales for requiring an in-class timed midterm examination in fall-semester small sections at Moritz is that it familiarizes students with law school exams. This includes exposure to analytical features such as issue spotting, discerning the relevance of facts, and applying rules in new situations. But the midterms also introduce students to the logistics of using exam software and anonymous identification, and alert them to the need for time management during the exam. Some professors regard familiarity with exam mechanics as one of the greatest benefits of the midterm exercise because it reduces anxiety and confusion during the final examination period. This potential “familiarity” effect of formative feedback deserves more attention in the law school setting where many students are faced with exam structures that they regard as a new experience.

2. *Voluntary nature of the exercise.* Professor Colker made this exercise voluntary, in part, because she did not want to waste her time grading answers that were not carefully done. She thought feedback would be useful only if students submitted what they considered to be their best work. She also wanted to spend as little time as possible providing feedback, given her busy schedule. Professor Colker knew that fewer students would take the practice exam question if it were voluntary. Would the association between taking the practice exam question and student grades have been as strong if the exercise were mandatory (but still ungraded)?

A mandatory exercise would eliminate any possible effects of selection associated with takers who are risk-averse, less overconfident, or proficient in study skills,<sup>126</sup> while providing the learning benefits of formative feedback to all the students in the class.

---

<sup>126</sup> See *supra* notes 105-106 and accompanying text.

However, aside from selection features associated with a voluntary exercise, it is also possible that the students who participated were those most likely to benefit from the experience due to their growth mindset.<sup>127</sup> If so, making the exercise mandatory might include students who would gain less from the experience. In other words, a rising tide of formative feedback might not lift all boats to the same extent. However, those disposed to take full advantage of the feedback opportunity would still gain, and even students with a fixed mindset might benefit to some degree from formative feedback. Moreover, there may be a “testing effect” aside from the effect of the feedback. The testing effect is a tendency for better performance on a final exam by students who took an initial test, even without feedback, relative to not taking a test.<sup>128</sup> Even if all students do not obtain a substantial benefit from the feedback, a mandatory practice exam would provide them all with a benefit due to any testing effect.

We suggest that the relationships among mindset,<sup>129</sup> the voluntary/mandatory design of formative feedback, and student performance warrant further measurement and analysis in the law school context.

3. *Ungraded nature of the exercise.* Because the exercise was optional, Professor Colker used it as an ungraded exercise in the sense that it did not count toward a taker’s course grade (although she did give takers a strong sense of the grade the answer would have earned).<sup>130</sup> Because the results of the exam were not factored into students’ grades,

---

<sup>127</sup> See *supra* text accompanying notes 109-114.

<sup>128</sup> Andrew C. Butler et al., *The Effect of Type and Timing of Feedback on Learning From Multiple-Choice Tests*, 13 J. EXPERIMENTAL PSYCHOL.: APPLIED 273, 274 (2007).

<sup>129</sup> Yan et al., *supra* note 113, at 141, measured mindset using questions developed by Chi-yue Chiu et al., *Lay dispositionism and implicit theories of personality*, 73 J. PERSONALITY & SOC. PSYCHOL. 19, 22 (1997).

<sup>130</sup> It is possible to design voluntary formative feedback that contributes to the final grade, see Nekby et al., *supra* note 102, at 354-55 (voluntary quizzes counted in the grade or could be taken as part of the final examination), but this might be difficult for a law school class.

Professor Colker could use a previous year's exam question and not worry about students cheating by getting outside help. This design also reduced her workload. If the question had counted in the final grade, she would have had to devise a new question and answer grid. She also would have had to be more careful in her estimate of the student's numerical grade, given its importance to a student's numerical average and the need to comply with Moritz's curve.

Would students have attained as much (or more) benefit from this exercise if the result had been factored into their grades? From a student perspective, the design of the exercise made it low risk and may have encouraged some students to participate who would not otherwise have done so. In contrast, the midterm exam that all Moritz students take in the fall is required, graded, and contributes a small proportion (5-10%) of the final grade in the small-section classes.<sup>131</sup> This policy was established on the theory that students would take the experience more seriously if it had some influence on their grade but that, given the small contribution, the exercise would still be low risk. On the other hand, some research suggests that when a student receives graded formative feedback, students focus on the grade rather than the comments.<sup>132</sup> However, if this is a danger, it would likely have been triggered by the estimated grade that Professor Colker communicated with the feedback. Her exercise could be considered as a hybrid of a graded and ungraded design.

4. *Timing of Feedback.* Students typically received written feedback within forty-eight hours of submitting the sample exam answer. Professor Colker could provide quick

---

<sup>131</sup> There is no way to measure the impact of the graded nature of these midterms because every student in a small section took the same graded midterm.

<sup>132</sup> See Sargent & Curcio, *supra* note 2, at 382 ("Numerous studies suggest that feedback may be more effective if ungraded because students tend to focus on grades, not suggestions for improvement.").

feedback because she gave students two weeks to complete the exam question (so answers were staggered) and because the exam was ungraded (so she didn't need to worry about cheating). If more students had taken the exam, and all the answers were submitted at the same time, she would have had difficulty providing such prompt feedback.

Would students have attained as much benefit if the feedback had been delayed? Law studies on formative assessment advocate prompt feedback<sup>133</sup> and cite research suggesting that feedback is more valuable if provided more quickly.<sup>134</sup> Research in other disciplines has yielded contradictory results.<sup>135</sup> Proponents of immediate feedback think that it is desirable because it enables learners to eliminate incorrect responses and reinforce correct responses.<sup>136</sup> Proponents of delayed feedback argue, in contrast, that it allows incorrect responses to dissipate so that they do not interfere with learning correct answers. Delaying feedback may also benefit learning due to the spaced presentation of the material, which improves the retention of information compared to a massed presentation (which would characterize immediate feedback).<sup>137</sup> There is empirical support for the spacing theory from laboratory experiments showing that delaying feedback boosted performance compared to immediate feedback.<sup>138</sup> This conclusion comes with a caution, however, that students may not fully process feedback after a delay

---

<sup>133</sup> See, e.g., Frost, *supra* note 112, at 946 (“Quick feedback is both responsive to students’ requests for faster turnaround and more effective than providing feedback after a long lag.”); Herbert Ramy, *Moving Students from Hearing and Forgetting to Doing and Understanding: A Manual for Assessment in Law School*, 41 CAP. U. L. REV. 837, 852 (2014) (“The first rule of effective feedback is that it must be prompt.”).

<sup>134</sup> See Sargent & Curcio, *supra* note 2, at 382; Schwarcz & Farganis, *supra* note 4, at 9.

<sup>135</sup> Butler et al., *supra* note 128, at 274, 280.

<sup>136</sup> *Id.* at 274.

<sup>137</sup> *Id.*

<sup>138</sup> *Id.* at 279-80 (finding that delayed feedback benefited both initially correct and incorrect responses).

unless they are required to do so, and thus may not engage in active processing in many applied settings.<sup>139</sup>

There is a problem, however, in applying the educational research conducted in other disciplines to law. “Delayed” feedback in these studies can mean as little as 10 minutes or one day after completing a question.<sup>140</sup> Thus many of the studies are testing effects of delay that are far shorter than the norm for law school feedback.<sup>141</sup> Professor Colker’s feedback, while delivered very quickly by law school standards, would be considered delayed in the terms of the educational studies.

At some point, delay is likely to become detrimental. It is certainly plausible that delaying feedback for too long will reduce students’ motivation to look at anything more than their grade.<sup>142</sup> After too long a delay, it becomes unrealistic to think that reviewing an exam will assist a student in adjusting her approach based on a long-past performance. For law school final exams, the typical months-long delay between taking a final exam and discussing it with the professor means that students’ memory of the exam (and the subject matter!) has faded and the relevance of the student’s responses is attenuated.<sup>143</sup> One fruitful area for research would be an examination of the effect of the length of delay on the effectiveness of feedback using time periods typical of law school conditions.

5. *Nature of the feedback.* Professor Colker provided both an estimated grade and written feedback on the practice exam answers. She also made herself available to meet with students for an oral discussion of the exam. Not only did she point out flaws in

---

<sup>139</sup> *Id.* at 280.

<sup>140</sup> *Id.* at 274.

<sup>141</sup> At Moritz, immediate feedback is not possible even for multiple-choice questions when the answers are machine graded. The answer sheets are processed at a central campus location with a turn-around time of several days.

<sup>142</sup> Butler et al., *supra* note 128, at 280.

<sup>143</sup> Ramy, *supra* note 133, at 853.

the student's analysis, often involving a poor use of the facts in the hypothetical, but she encouraged weak-performing students to re-write their answer and return for an additional conference. She used Word's "comment" feature so she could make comments directly tied to a particular sentence or paragraph in the student's answer.

The literature suggests that feedback is more effective if it provides an explanation rather than merely indicating a correct response.<sup>144</sup> Students benefit more when the feedback provides specific information about what they did, and did not, do well.<sup>145</sup> This specificity better allows students to adjust their approach based on the professor's comments.<sup>146</sup> Professors could improve their comments by including specific indications of the weaknesses and strengths of the answer. One example of this type of improvement, offered by Ramy, is "writing 'weak analysis—facts missing that were relevant to defendant's intent' as opposed to merely writing 'weak analysis.'"<sup>147</sup>

Effective feedback depends on both the experience that the professor provides to the students and the way that students receive or interpret it.<sup>148</sup> For a formative assessment to help a student develop her knowledge or skills, a student must (1) understand the goal or standard that is the aim of the assessment; (2) compare her performance to that goal or standard; and (3) take appropriate steps to close the gap between her performance and the goal.<sup>149</sup> Thus, the ability and inclination of students to

---

<sup>144</sup> See Sargent & Curcio, *supra* note 2, at 381, 400.

<sup>145</sup> See Schwarcz & Farganis, *supra* note 4, at 9.

<sup>146</sup> Ramy, *supra* note 133, at 853. Specificity can be provided by comments from the professor, peer scoring, or grading rubrics. *Id.* at 854.

<sup>147</sup> *Id.*; see also Paula J. Manning, *Understanding the Impact of Inadequate Feedback: A Means to Reduce Law Student Psychological Distress, Increase Motivation, and Improve Learning Outcomes*, 43 CUMB. L. REV. 225, 246-55 (2013) (critiquing common feedback patterns and suggesting improvements).

<sup>148</sup> Sargent & Curcio, *supra* note 2, at 381.

<sup>149</sup> FEEDBACK FOR LEARNING 21 (Susan Askew ed., 2000); D. Royce Sadler, *Formative Assessment and the Design of Instructional Systems*, 18 INSTRUCTIONAL SCI. 119, 121 (1989).

use feedback are important considerations if the goal is to improve the quality and effectiveness of formative assessment.

Further work is needed to determine what types of feedback are most effective for law students, whether students differ in their feedback needs, and whether those needs change over the course of a legal education. Existing studies suggest that some students benefit more from feedback than others—but those studies point in different directions. While Curcio’s two studies found a stronger relationship between formative assessments and grades for students with higher LSAT scores and UGPAs,<sup>150</sup> Schwarcz and Farganis found the opposite: In their equations, the association was strongest for students with below-median entering credentials.<sup>151</sup> Our results were most similar to those of Schwarcz and Farganis; we found that the coefficients for feedback were largest for students with below-median entering credentials.<sup>152</sup> On the other hand, the differences we identified were not as marked as the ones Schwarcz and Farganis found; overall, students who took Professor Colker’s practice exam obtained higher grades regardless of their pre-law credentials.

These varied results might stem from different types of feedback. The feedback administered by Professor Curcio, for example, might have been especially useful to students with above-median entering credentials.<sup>153</sup> The professors studied by Schwarcz

---

<sup>150</sup> *Does Practice Make Perfect?*, *supra* note 8, at 293-98 (LSATs and UGPAs above the median); Sargent & Curcio, *supra* note 2, at 391 (LSATs and UGPAs for the highest 2/3 of the class).

<sup>151</sup> Schwarcz & Farganis, *supra* note 4, at 30.

<sup>152</sup> See *supra* notes 81-85 and accompanying text; tbls. 7-10; note 94 and accompanying text.

<sup>153</sup> In the first Curcio study, the backbone of the feedback was provided as annotated model answers that students used for a self- or peer-edit. This method was supplemented for single assessments with oral class feedback, a grading rubric, and one paper graded with comments from the professor. *Does Practice Make Perfect?*, *supra* note 8, at 287-89. The second Curcio study again used annotated model answers, but added grading rubrics and self-reflective exercises. Sargent & Curcio, *supra* note 2, at 386-88. The increased emphasis on student self-reflection was designed to ensure that students received the feedback mindfully. *Id.* at 386.



and Farganis, in contrast, might have provided feedback that was more appropriate for below-median students.<sup>154</sup> Future work should look more closely at the precise nature of feedback offered students so that we can identify any relationships between student credentials and the most effective forms of feedback.<sup>155</sup>

### *C. Formative Feedback and Grades in Other Courses*

With respect to performance in classes other than Constitutional Law, the results showing a relationship between taking the practice exam question and grades in those classes are intriguing. One could imagine that the time used to study for and complete the Constitutional Law practice exam question would have taken away time that students spent on their other classes. Thus, although it was easy to predict that there would be an association between taking the practice exam question and student performance in Constitutional Law, it was not clear that the association would extend to other classes. Yet, our results suggest that there was an association between taking the practice exam question and performance in other classes.

Especially when combined with the results reported by Schwarcz and Farganis, our findings lend support to the theory that formative feedback promotes academic growth, and that this growth translates to subjects other than the one in which a student

---

<sup>154</sup> The professors in that study offered feedback in a range of forms, including grades on multiple choice tests, graded essay exams and assignments, written comments, and oral feedback individually and in small groups. Schwarcz & Farganis, *supra* note 4, at 17-18. The study did not consider individualized feedback to include model answers, grading rubrics, or in-class comments on strong answers or common mistakes. *Id.* at 17. As noted above, other factors may also explain the different outcomes in these studies. *See supra* notes 57-60 and accompanying text.

<sup>155</sup> Curcio and her co-authors, for example, suggest that some types of formative assessment (like annotated model answers) may be most effective for students with strong metacognitive skills. *Does Practice Make Perfect?*, *supra* note 8, at 302. Students who lack those skills may “lack the ability to distinguish between the standard exemplified by the model answer and their own work.” Frost, *supra* note 112, at 949. Other, more individualized forms of feedback may offer particular benefit to students with weak metacognitive skills.

received feedback. Schwarcz and Farganis eliminated selection bias from their study, because their subjects did not choose whether to receive formative feedback. On the other hand, they could not rule out the possibility that professor differences apart from feedback promoted student achievement. Our study allowed us to rule out the latter differences, although we could only partially control for selection bias. The fact that these complementary studies produced similar results increases the plausibility of a causal relationship between formative feedback and student achievement—although, of course, further work needs to be done to confirm this hypothesis.

Assuming that the formative feedback in our study was causally linked to higher grades in other courses, several paths could explain that result:

1. The feedback from the Constitutional Law exam could have helped students improve their exam-taking skills, analytic skills, and writing in a manner that enhanced performance in all classes, even though the assessment instruments differed. Assessments in other courses ranged from memos written outside of class (LAW II) to time-pressured final exams with a combination of essay and multiple-choice questions. Yet our analysis showed a relationship between taking the practice exam and superior performance on a range of assessments.<sup>156</sup>

2. Taking time to review Constitutional Law in the middle of the semester might have permitted students to stay on top of that class, reducing the time required for review at the end of the semester. That, in turn, could have freed up time at the end of the semester to study for exams in other classes.

---

<sup>156</sup> See *supra* notes 87- 89 and accompanying text. Note, in particular, that students who took the practice exam achieved significantly higher grades in both LAW II (a writing course) and Contracts (a course with a traditional timed exam).

3. Taking the time to review Constitutional Law in the middle of the semester could have encouraged students to develop good study skills in all their classes. It also might have prompted them to stay on top of all classes rather than wait to cram at the end.

4. Students who took the practice exam might have had a growth mindset and used that growth mindset to improve their performance in all their classes. While Professor Colker may have been the only professor to offer a practice exam, other professors probably held office hours, reviewed hypotheticals in class, and engaged in other kinds of work that provided opportunities for formative feedback. The students who completed Professor Colker's practice exam question may have been more likely than their classmates to pursue those opportunities in other courses.

One limit on this hypothesis is the fact that students who took the practice exam did not achieve significantly higher fall-semester grades than their classmates. If the former group of students enjoyed a growth mindset, that orientation should have helped them in both the fall and spring. On the other hand, law school may be so challenging that the benefits of a growth mindset do not appear until after students adjust to the new environment.

Alternatively, *choosing* to pursue a feedback opportunity may be necessary to trigger a growth mindset. Under this theory, Moritz's fall semester—with mandatory feedback offered to all students—was not enough to activate the benefits of a growth mindset. Instead, students capitalized on their growth mindset only when formative feedback became optional.

#### D. Gender

The gender factor deserves further examination. In all three years, Professor Colker was the only female professor that these students had in the spring semester for a podium class.<sup>157</sup> Gender was not a significant factor predicting grades in her class, but it was significantly associated with grades in at least two other spring-semester classes.<sup>158</sup> Notably, male students obtained significantly higher grades than women in Legislation, a course taught exclusively by male professors during this period. Conversely, women fared significantly better than men in LAW II, a course taught predominately by female professors.

Surprisingly little research has examined the interactions among student gender, professor gender, and course grades in college or graduate programs.<sup>159</sup> One laboratory study found that both men and women achieved higher quiz grades after viewing a short lecture by a male professor than by a female one.<sup>160</sup> The authors hypothesized that “students [may] pay less attention to, or put less credence on, the lecture of a non-normative professor than to the more normative one.”<sup>161</sup> That study, however, used highly artificial conditions to study achievement.<sup>162</sup> Our findings, drawn from full-

---

<sup>157</sup> These students were exposed to female professors during the fall semester to varying degrees. In 2013-14, Professor Colker’s students had a female professor for Torts in the fall and half of them had a female professor for Criminal Law in the fall. In 2014-15, Professor Colker’s students had a female professor for Torts in the fall. In 2015-16, Professor Colker’s students had female professors for Torts and Civil Procedure in the fall. In addition, many of their LAW I and LAW II (legal analysis and writing) classes were taught by females.

<sup>158</sup> See *supra* notes 98-99 and accompanying text.

<sup>159</sup> See Basow et al., *supra* note 115, at 354-55.

<sup>160</sup> *Id.* at 358.

<sup>161</sup> *Id.* at 361.

<sup>162</sup> The study used a computer-animated talking head to deliver the lecture; the lecture lasted only three minutes; and achievement was measured through a ten-question true/false quiz delivered immediately after the lecture. *Id.* at 356-57.

semester classes, suggest that law students may achieve higher grades with same-gender professors than opposite-gender ones.<sup>163</sup>

On the other hand, the courses in which we identified gender differences also differed markedly in assessment methods. Professors in the Legislation course, where male students outscored female ones, for the most part evaluated students with a single, end-of-semester, timed exam.<sup>164</sup> In LAW II, where women outperformed men, grades rested on a series of writing assignments completed outside of class. Some research suggests that men perform better than women on time-pressured exams.<sup>165</sup> Women, in contrast, may achieve higher scores on out-of-class writing assignments.<sup>166</sup> Our findings may reflect those differences.

Whatever the origin of the gender differences in our study, there is no doubt that those discrepancies exist. Women were significantly more likely than men to take an optional practice exam; they achieved significantly higher grades in a legal writing course; and they registered significantly lower grades in at least one podium course. Law

---

<sup>163</sup> Cf. Schroeder & Mynatt, *supra* note 116 (finding some disadvantages to graduate students, especially women, with opposite-gender advisors).

<sup>164</sup> In one year, the Legislation professor used a hybrid two-hour in-class timed exam and five-hour take-home exam.

<sup>165</sup> See Maria De Paola & Francesca Gioia, *Who Performs Better under Time Pressure?: Results from a Field Experiment*, IZA Discussion Paper No. 8708 (December 2014) (finding that being exposed to time pressure exerts a negative and statistically significant impact on students' performance on both verbal and numerical tasks with a significant negative gender difference (adverse to women) in the impact of performance on verbal tasks); William C. Kidder, *Portia Denied: Unmasking Gender Bias on the LSAT and its Relationship to Racial Diversity in Legal Education*, 12 YALE J. LAW & FEMINISM 1, 11 (2000) (noting that "women were disadvantaged relative to men by the LSAT in each of the sixteen UGPA bands," suggesting that the speeded nature of the LSAT disadvantages women in the law school admission cycle in comparison to men with the same UGPA).

<sup>166</sup> Cf. William D. Henderson, *The LSAT, Law School Exams, and Meritocracy: The Surprising and Undertheorized Role of Test-Taking Speed*, 82 TEX. L. REV. 975, 1043-44 (2003-2004) (reporting that the LSAT is a less useful predictor of grades in courses with take-home exams or other untimed writing exercises). While Henderson did not consider gender in his study, it does suggest that different factors predict performance in classes with writing assignments than classes with timed in-class exams. Our study suggests that gender might be one additional factor that might help predict performance in a class with an untimed writing assignment.

schools should examine traditional pedagogies to determine whether those approaches advantage one gender over the other.

#### *E. Race*

We found that race was not significant in predicting whether a student would take Professor Colker's practice exam or in predicting the grade in her course, after controlling for other variables. Nor did race show any association with grades in other spring-semester courses, once we controlled for those variables. Nor, finally, did race affect the positive association between taking the practice exam and course outcomes. The association between that exam and grades (in both Professor Colker's course and other spring-semester courses) was equally strong for white and nonwhite students.

All of our equations, however, controlled for a student's fall-semester GPA. This control masked a race effect that numerous other studies have identified. If we removed fall-semester GPA from our regression equation, the variable for race assumed significance: nonwhite students received significantly lower grades than their white classmates in spring-semester classes.<sup>167</sup> The same effect occurred in the fall.<sup>168</sup> These analyses confirmed a consistent and disturbing finding: nonwhite law students secure significantly lower grades than white students, even after controlling for entering credentials.<sup>169</sup> Law schools must work harder to identify the forces producing this discrepancy.

On the other hand, our study produced one promising lead: when we examined students with LSAT scores at or above their class median, nonwhite students achieved

---

<sup>167</sup> This analysis is available from the authors upon request.

<sup>168</sup> This analysis is available from the authors upon request.

<sup>169</sup> See *supra* note 101 and accompanying text.

significantly higher grades in Professor Colker's class (after controlling for entering credentials and fall-semester GPA) than white students. We reiterate the need for caution in drawing conclusions from this result: the number of students in our non-white group was small, and the small number required us to combine students from multiple minority groups. Still, the results suggest that—with further study—law schools may be able to identify the circumstances that nurture success in nonwhite students.

## V. CONCLUSION

Our study generated several notable findings about formative assessment in legal education, using data from Professor Colker's voluntary practice exam question in a required first-year Constitutional Law course. First, in a setting where the assessment was voluntary, female students and students with higher UGPAs disproportionately sought the feedback. Second, obtaining formative feedback in Constitutional Law was associated with a higher grade in that class. That relationship was both statistically and practically significant. Third, completing this formative assessment was associated with higher grades in all of the students' *other* spring semester classes; that relationship reached statistical significance for two of those classes. When we controlled for LSAT, UGPA, and other variables, moreover, taking the practice exam retained its significant association with grades in other spring-semester classes.

Our study, finally, identified several intriguing gender and race differences. After controlling for academic credentials, men received significantly higher grades than women in one spring-semester class; women, conversely, outperformed men in a second course. After controlling for entering credentials, meanwhile, nonwhite students obtained

significantly lower grades than white students, an effect that several other studies have noted. We, however, also found that nonwhite students with at-or-above median LSAT scores achieved higher grades than white students in Professor Colker's Constitutional Law course. All of these effects deserve further study.

We will use our larger data set to explore some of the questions raised here; we hope that professors at other schools will also pursue experiments shedding further light on formative feedback, gender differences, and race effects. Based on our work, we suggest that one fertile avenue for investigation might be the relationship between students' mindsets and ways to maximize the benefit of formative feedback. Our study and the Minnesota one also suggest that formative assessment in one course can improve performance in other courses taught the same semester. Researchers, therefore, should document the full range of formative feedback offered students and should search for impacts beyond the course in which the feedback was offered.