

The Impact of Individualized Feedback on Law Student Performance

Daniel Schwarcz and Dion Farganis

Overview

For well over a century, first-year law students have typically not received any individualized feedback in their core “doctrinal” classes other than their grades on final exams. Although critics have long assailed this pedagogical model, remarkably limited empirical evidence exists regarding the extent to which increased feedback improves law students’ outcomes. This article helps fill this gap by focusing on a natural experiment at the University of Minnesota Law School. The natural experiment arises from the assignment of first-year law students to one of several “sections,” each of which is taught by a common slate of professors. A random subset of these professors provides students with individualized feedback other than their final grades. Meanwhile, students in two different sections are occasionally grouped together in a “double-section” first-year class. We find that in these double-section classes, students in sections that have previously or concurrently had a professor who provides individualized feedback consistently outperform students in sections that have not received any such feedback. The effect is both statistically significant and hardly trivial in magnitude, approaching about one-third of a grade increment after controlling for students’ LSAT scores, undergraduate GPA, gender, race, and country of birth. This effect corresponds to a 3.7-point increase in students’ LSAT scores in our model. Additionally, the positive impact of feedback is stronger among students whose combined LSAT score and undergraduate GPA fall below the median at the University of Minnesota Law School. These findings substantially advance the literature on law school

Daniel Schwarcz is a Professor at the University of Minnesota Law School. **Dion Farganis** is a 2017 graduate of the University of Minnesota Law School, current judicial law clerk to the Honorable Joan N. Ericksen of the U.S. District Court for the District of Minnesota, and former Assistant Professor of Political Science at Elon University. For helpful comments, we thank Robert Ahdieh, Michael Asimow, June Carbone, Bradley Clary, Andrea Curcio, Allan Erbsen, Jay Feinman, Dan Ho, Jason Husser, Bert Kritzer, William McGeeveran, Noelle Noonan, Kathleen O’Neill, Hari Osofsky, Tamar Resnick, Richard Sander, Michael Simkovic, Sandra Simpson, and an anonymous reviewer, as well as participants in faculty workshops at University of Minnesota, Southwestern Law School, St. Thomas Law School, and UCLA Law School. We also acknowledge the wonderful support provided by the University of Minnesota Library generally, and Connie Lenz in particular. Luke Wolf provided excellent research assistance.

pedagogy, demonstrating that individualized feedback in a single class during the first year of law school can improve law students' exam quality in all their other classes. In light of the broader literature on the importance of formative feedback in effective teaching, these findings suggest that, at a minimum, law schools should systematically provide first-year law students with individualized feedback in at least one "core" doctrinal first-year class before final exams. Doing so would almost certainly have positive distributional consequences and improve the fairness of law school grades. It would also likely promote students' acquisition of relevant legal skills. Finally, this reform would help implement the American Bar Association's recent requirement that law schools utilize formative assessment methods in their curricula.

Introduction

For well over a century, students' grades in most law school classes have been based exclusively on their performance on a single end-of-semester exam.¹ Many weeks after these exams are complete and classes are concluded, law students typically receive a single piece of feedback consisting of a letter grade.² Students generally do not receive any individualized comments regarding their exam performance at this time.³ And at no point before receiving their final grades do students in most classes receive any specific and individualized feedback on their understanding and mastery of the material.⁴

For almost as long as this educational model has been in place, critics have emphasized its pedagogical deficiencies.⁵ Effective education, these critics suggest, requires "frequent formative assessments that provide students with the opportunity to gauge their progress as they acquire new skills."⁶ Ideally, such feedback should be promptly provided to students, at a time when they remain immersed in the underlying material and capable of adjusting their approach.⁷ This feedback, moreover, should specifically identify the strengths

1. See Phillip C. Kissam, Essay, *Law School Examinations*, 42 VAND. L. REV. 433, 471-72 (1989); Robert C. Downs & Nancy Levit, *If It Can't Be Lake Woebe gone . . . A Nationwide Survey of Law School Grading and Grade Normalization Practices*, 65 UMKC L. REV. 819, 822-23 (1997).
2. See Downs & Levit, *supra* note 1, at 823.
3. Some professors provide students with a grading rubric for the final exam or model student answers. Many others offer to meet individually with students who proactively request meetings to discuss their exams.
4. See Downs & Levit, *supra* note 1, at 823; GREGORY S. MUNRO, *OUTCOMES ASSESSMENT FOR LAW SCHOOLS* 34 (2000).
5. For early criticisms, see, e.g., Ben D. Wood, *The Measurement of Law School Work*, 24 COLUM. L. REV. 224 (1924).
6. Herbert N. Ramy, *Moving Students from Hearing and Forgetting to Doing and Understanding: A Manual for Assessment in Law School*, 41 CAP. U. L. REV. 837, 837 (2013).
7. See *id.* at 852-53; Anthony Niedwiecki, *Teaching for Lifelong Learning: Improving the Metacognitive Skills of Law Students Through More Effective Formative Assessment Techniques*, 40 CAP. U. L. REV. 149, 178 (2012).

and weaknesses of students' performance to focus their efforts.⁸ The single end-of-semester law school exam fails along each of these dimensions: Feedback is provided at a single point in time, well after students have completed the class, without any specificity about what the student did and did not do well.⁹ As a result, final exams in law school operate predominantly to sort students into different categories, but do little to promote learning.¹⁰

Over the past decade, as law schools have faced a well-documented decline in applications and increase in competitive pressures,¹¹ these critiques of traditional law school education have reached a fever pitch. Many articles have criticized the traditional model of law school pedagogy,¹² some schools have implemented mandatory midterm exams in first-year classes,¹³ and numerous individual professors have integrated different forms of feedback into their classes. At the same time, the American Bar Association has recently adopted new standards that require law schools to articulate specific learning outcomes and to "utilize both formative and summative assessment methods in [their] curricul[a] to measure and improve student learning and provide meaningful feedback to students."¹⁴

Despite the emerging orthodoxy that traditional methods of legal education are deeply flawed, remarkably limited empirical evidence exists of the extent to which better and more frequent feedback can actually improve law students' performance.¹⁵ This lack of empirical evidence has provided cover for both instructors and law schools that have been slow to incorporate individualized feedback into law school classes. After all, providing meaningful and prompt feedback to students can be difficult and time-consuming, causing many within the legal academy either to ignore or dismiss the well-worn criticisms of the standard law school educational model.¹⁶

8. Rogelio A. Lasso, *Is Our Students Learning? Using Assessments to Measure and Improve Law School Learning and Performance*, 15 BARRY L. REV. 73, 75 (2010).
9. See, e.g., Kissam, *supra* note 1, at 441-46; MUNRO, *supra* note 4, 35-36, 151; Downs & Levit, *supra* note 1, at 822-23.
10. Janet Motley, *A Foolish Consistency: The Law School Exam*, 10 NOVA L. J. 723, 723-24 (1986).
11. See, e.g., William D. Henderson, *A Blueprint for Change*, 40 PEPP. L. REV. 461 (2013).
12. See, e.g., Olympia Duhart, "It's Not for a Grade": *The Rewards and Risks of Low-Risk Assessment in the High-Stakes Law School Classroom*, 7 ELON L. REV. 491 (2015); Elizabeth M. Bloom, *A Law School Game Changer: (Trans)formative Feedback*, 41 OHIO N.U. L. REV. 227 (2015); 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 (2013).
13. See survey conducted by Anahid Gharakhanian, Vice Dean, Southwestern Law School (Jan. 2016) (on file with author) (reporting in informal survey of associate deans that several law schools require first-semester midterms, although most do not and instead leave the issue up to individual instructors' discretion).
14. See Am. Bar Ass'n Standard 314 (2015).
15. See *infra* Part II (reviewing the empirical literature).
16. *But cf.* Ramy, *supra* note 6, at 854 (suggesting various approaches to integrating formative feedback into the classroom without placing undue burdens on instructors of large classes).

This article seeks to address this gap in the literature by supplying and analyzing some empirical evidence about the impact that individualized feedback can have on law students' development. To do so, the article exploits a natural experiment at the University of Minnesota Law School involving first-year law students.¹⁷ Like most law schools, first-year law students at Minnesota are assigned to one of several individual sections upon starting law school. These sections consist of a cohort of between forty and fifty first-year students, each of which is assigned a common set of professors to teach the six required doctrinal first-year classes: contracts, civil procedure, torts, and constitutional law in the first semester, and property and criminal law in the second semester. Due to shifting staffing considerations, however, students in two different sections are occasionally grouped together in a single "double-section" class consisting of between eighty and a hundred students.

Because the individual instructors who teach within the first-year curriculum at the law school vary significantly in whether and how they deliver feedback before or in addition to the standard end-of-semester law school exam, students from two sections that are grouped together in a double-section class have often received different levels of feedback in their single-section classes. For instance, students in Section A might have a contract law professor who provides prompt individualized feedback before the final exam, whereas students in Section B might not have any professors who provide such feedback. Students from Section A and Section B might then be combined into a double-section constitutional law class in which they are blindly graded on the same curve by the same professor on the same exam. If individualized feedback has a positive impact on law students beyond the class in which it is received, then one might hypothesize that students in Section A, who had a contract law professor who provided individualized feedback, would outperform students in Section B, who did not receive any individualized feedback, in their common double-section constitutional law class.

This, in fact, is exactly what we observe. The effect is both statistically significant and hardly trivial in magnitude, even after controlling for students' LSAT scores, undergraduate GPA, gender, race, and country of birth. Among the eight double sections at Minnesota between fall 2011 and fall 2015 in which one section received individualized feedback in one of its classes before the double-section final exam and the other did not, the students from the section receiving individualized feedback outperformed the students from the section that did not *in every single class*. The likelihood of this occurring by chance is one in 256. The magnitude of these differences varied from razor-

17. The article's basic research design is most similar to Daniel E. Ho and Mark G. Kelman, *Does Class Size Affect the Gender Gap? A Natural Experiment in Law*, 43 J. LEGAL STUD. 291 (2014), which leveraged the random assignment of law students into sections to assess the impact of educational reforms. For further discussion of Ho and Kelman, see *infra* note 38 and accompanying paragraph. We characterize this study as exploiting a "natural experiment" because the students who received individual feedback were determined by factors outside of the authors' control (by the registrar's assignment of students to sections) in a manner that was largely random.

thin to more than half a grade increment (.235 on a 4.0 GPA scale), with the average difference across all eight double sections coming in at approximately .12 on a 4.0 scale. When the data from all eight double sections are combined and students' individual characteristics are controlled for through regression techniques, the effect is statistically significant at a ninety-nine percent level of confidence ($p = .010$). The impact of individualized feedback on students' grades is equivalent to a 3.7-point increase in students' LSAT scores in our model. Further analysis limits the possibility that this effect is driven either by the relative clarity in presentation of professors who provide individualized feedback or the possible tendency of these professors to place fewer demands on their students during final exam periods. It also suggests that the positive impact of feedback is concentrated among students whose combined LSAT score and undergraduate GPA fall below the median at the University of Minnesota Law School.

Unlike prior research, these results do not simply suggest that individualized feedback improves students' performance in the class where such feedback is given. Instead, they suggest that individualized feedback in a single first-year doctrinal class can improve the quality of students' exams *in all other* traditional law school classes during the first year of law school. This finding has a variety of important implications. Most directly, it suggests that the accuracy and fairness of law school grades can be compromised when law schools do not provide students with consistent levels of individualized feedback. More importantly, our results suggest the possibility that providing students with enhanced individualized feedback can promote their acquisition of the types of legal skills that are tested on standard law school exams, such as issue-spotting, applying relevant legal standards to complex factual settings, and analyzing policy. The evidence that individualized feedback appears to have a stronger effect on below-median students also has important implications. In particular, it indicates that individualized feedback can disproportionately benefit the subset of students who incur the largest costs to attend law school and who are most at risk of failing the bar exam or being unable to land desirable postgraduate employment.

To be sure, our results leave open several important questions that should be studied in future research. First, data limitations preclude us from offering any robust evidence of the forms of individualized feedback that are most effective. Instead, we simply define individualized feedback based on our reading of the broader literature on effective pedagogy. This definition of individualized feedback includes instructor-provided written or oral comments on individual students' exams or exam-like assignments, as well as multiple-choice exams that count toward students' ultimate grades and are designed to test higher-order skills (such as applying rules to facts or analyzing hypotheticals). By contrast, it does not include generalized in-class instructor discussion of mock exams or assigned problems, nor does it include any feedback on "practical" writing assignments, such as contract- or complaint-drafting exercises. Of course, it is almost certain that some types of individualized feedback are

more effective than others. Interestingly, this reality actually suggests that the most effective forms of individualized feedback—whatever they turn out to be¹⁸—likely have a larger effect on student performance than what we measure here.¹⁹ This is because our results measure only the average impact of all forms of individualized feedback, including both those types of feedback that are relatively effective and those that are comparatively ineffective.

Second, our results leave unaddressed the important question of the mechanism by which individualized feedback improves student performance. Individualized feedback may improve students' performance on exams by developing important legal skills, such as the ability to spot relevant legal issues or apply the law to a complicated fact pattern. Alternatively, such feedback may simply improve students' ability to "game" law school exams without developing the legal skills that law school exams endeavor to test. Yet another possibility is that individualized feedback may prompt students to study more effectively or boost their confidence. The mechanism by which individualized feedback has the impact we identify is relevant to a broad range of related issues, such as the desirability of providing individualized feedback after the first year of law school, the marginal impact of providing individualized feedback in more than a single class in the first year of law school, and the extent to which individualized feedback has a long-term benefit for students in terms of metrics such as bar passage and employment outcomes.

Despite these limitations, we believe that, on balance, our findings have a clear normative implication: All law schools should, at a minimum, systematically provide first-year law students with individualized feedback on their performance on an exam or exam-like assignment in at least one "core" doctrinal first-year class before final exams. This suggestion is intentionally conservative, reflecting the limitations of our findings as well as the obvious cost, in terms of time and productivity, to professors who provide individualized feedback. Most law schools and law professors, we believe, should aspire to go well beyond this limited policy proposal in incorporating enhanced feedback into their teaching. To be sure, this broader recommendation goes beyond our data and implicates a variety of complicated tradeoffs. But it is at least grounded in our findings, which do indeed suggest that there is limited reason to believe that professors' prevailing practice of largely forgoing individualized feedback in law school classes reflects an informed trade-off between the goal of training effective lawyers and competing goals, such as promoting influential legal scholarship.

Our analysis proceeds in six parts. Part I reviews the extant literature on individualized feedback and law school pedagogy, providing a detailed critique of the limited empirical evidence that is specific to the law school setting. Part II describes the natural experiment that undergirds our analysis, while Part

18. We hypothesize that the most effective form of individualized feedback consists of written feedback on prefinal exams that count for a small, but meaningful, percentage of students' final grades.

19. We thank Richard Sander for making this point to us.

III reviews the data we utilize. In Part IV we present our results. Part V then addresses two of the most important objections to our empirical strategy. Finally, Part VI discusses what we take to be the normative implications of our findings. For readers interested in the bottom-line implications of our conclusions, Part VI may be the most relevant and interesting section of the article.

I. The Extant Literature on Feedback and Law School Pedagogy

A robust body of empirical literature demonstrates that feedback can play a vital role in promoting learning, both in higher education and in a variety of other settings. Dozens of law review articles and books rely on this literature to criticize the standard law school pedagogical model and promote change. Yet remarkably limited empirical evidence demonstrates that better feedback can improve students' performance in the law school setting. This section critically reviews the extant literature on these topics. Part A canvasses the predominant approach in the literature on law school pedagogy, which seeks to apply the generalized literature on education to the law school setting without empirically testing whether or how this approach would improve law students' learning outcomes. Part B then critiques the small body of empirical literature that does examine the impact that providing feedback has on law school performance.

A. Extensive Theoretical Literature

There is no shortage of law review articles bemoaning the traditional law school pedagogical approach generally, and the lack of feedback in law school classes in particular.²⁰ Although these articles obviously vary substantially in their structure and central themes, they often emphasize three key points: (a) the importance of "formative feedback" to learning; (b) the lack of formative feedback in traditional law school classes; and (c) the various possibilities for providing students with formative feedback in law school.

Starting with the first of these themes, most of the literature on promoting better law school pedagogy reviews empirical educational research conducted outside of law schools suggesting the importance of formative feedback.

20. See, e.g., Duhart, *supra* note 12 (reviewing both the risks and rewards of providing students with formative feedback throughout the semester); Bloom, *supra* note 12 (providing advice on developing formative feedback techniques that are likely to succeed); E. Scott Fruehwald, *How to Help Students from Disadvantaged Backgrounds Succeed in Law School*, TEX. A&M L. REV., Fall 2013, 83, 115-17 (suggesting that better formative assessment, when combined with techniques for developing metacognition, can improve disadvantaged students' performance in law school); Niedwiecki, *supra* note 7, at 180-85 (encouraging professors to make greater use of self-assessment tools in the process of providing formative assessment); Lasso, *supra* note 8, at 89 (advocating for the provision of prompt and frequent feedback to law students); Steven Friedland, *A Critical Inquiry into the Traditional Uses of Law School Evaluation*, 23 PACE L. REV. 147, 206-10 (2002) (encouraging the use of assessment as an educational tool); Steve Sheppard, *An Informal History of How Law Schools Evaluate Students, with a Predictable Emphasis on Law School Final Exams*, 65 UMKC L. REV. 657 (1997); Jay M. Feinman & Marc Feldman, *Achieving Excellence: Mastery Learning in Legal Education*, 35 J. LEGAL EDUC. 528 (1985).

This research demonstrates both that formative feedback is vital to effective education generally²¹ and to higher education in particular.²² In most of these studies, formative feedback is defined as feedback provided to students with the primary goal of helping them to learn by providing “a clearer picture of areas that need improvement.”²³ It is typically contrasted with summative feedback, which is designed primarily to assess whether students have met specified learning objectives.²⁴ Although instructors can use multiple techniques to provide students with formative feedback, the research suggests that such feedback is generally most effective when it is prompt and specific about the good and bad elements of students’ work.²⁵

After emphasizing the importance of formative feedback to effective education, the literature on law school pedagogy almost uniformly criticizes the traditional law school model of assessing student performance based on a single end-of-semester exam. This practice, the literature suggests, provides students principally with summative, rather than formative, feedback.²⁶ In part, this is because students are informed only about how they performed weeks or months after the exams have been taken, once the course is completed.²⁷ Grades on traditional law school exams also do not generally provide students with any specific information about what they did or did not do well on the exam.²⁸ The result, many commentators have suggested, is not simply that students learn less well, but also that they are subject to undue amounts of uncertainty and anxiety throughout their first year of law school.²⁹

21. See, e.g., COMM. ON DEVS. IN THE SCI. OF LEARNING, HOW PEOPLE LEARN: BRAIN, MIND, EXPERIENCE, AND SCHOOL 24–25 (John D. Bransford et al. eds., expanded ed. 2000); David J. Nicol & Debra Macfarlane-Dick, *Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice*, 31 *STUD. HIGHER EDUC.* 199, 200 (2006); Michael Filsecker & Michael Kerres, *Repositioning Formative Assessment from an Educational Assessment Perspective: A Response to Dunn & Mulvenon* (2009), *PAC. ASSESSMENT, RES. & EVALUATION*, Dec. 2012, at 1–2; Dylan Wiliam, *The Role of Formative Assessment in Effective Learning Environments*, in *THE NATURE OF LEARNING: USING RESEARCH TO INSPIRE PRACTICE* 135 (H. Dumont et al. eds., 2010); Valerie J. Shute, *Focus on Formative Feedback*, 78 *REV. EDUC. RES.* 153 (2008).
22. See, e.g., M.I. Núñez-Peña, R. Bono & M. Suárez-Pellicioni, *Feedback on Students’ Performance: A Possible Way of Reducing the Negative Effect of Math Anxiety in Higher Education*, 70 *INT’L J. EDUC. RES.* 80 (2015); Carol Evans, *Making Sense of Assessment Feedback in Higher Education*, 83 *REV. EDUC. RES.* 70, 73 (2013).
23. Ramy, *supra* note 6, at 845.
24. See Duhart, *supra* note 12, at 496–98.
25. MUNRO, *supra* note 4, at 151; see also Manning, *supra* note 12, at 257.
26. See, e.g., Steve H. Nickles, *Examining and Grading in American Law Schools*, 30 *ARK. L. REV.* 411 (1977); Friedland, *supra* note 20, at 166–67; Michael Hunter Schwartz, *Teaching Law by Design: How Learning Theory and Instructional Design Can Inform and Reform Law Teaching*, 38 *SAN DIEGO L. REV.* 347, 408–09 (2001); Sheppard, *supra* note 20, at 681; Kissam, *supra* note 1, at 451–52.
27. Sheppard, *supra* note 20, at 681.
28. See Ramy, *supra* note 6, at 837.
29. Manning, *supra* note 12, at 227–29.

The extant literature also offers law school instructors numerous potential approaches to providing students with more formative feedback, often in a manner that can be implemented even in relatively large classes. For instance, the literature suggests that multiple-choice exams can be an effective technique for providing formative assessment, particularly because the feedback can be provided to students promptly.³⁰ Additional techniques for providing meaningful formative assessment to large groups of students include the use of grading rubrics, group feedback, peer feedback, and self-assessment tools.³¹

B. Limited Empirical Evidence

Despite the extensive literature criticizing the lack of formative feedback in law school classes, the empirical evidence evaluating the impact of this feedback is remarkably limited. The two most notable empirical studies of feedback in law school were conducted in 2008 and 2012, by Curcio and her co-authors. The 2008 study examined the impact of providing enhanced feedback to students in a single eighty-person first-year civil procedure class.³² Students in Curcio's class were required to write and hand in five three-page essays over the course of the semester, and each student received individualized feedback from the professor teaching the class on one of the first four of these assignments. These students were then given the same final exam as students in a different first-year section of civil procedure. This control-group class took place in the same year and at the same law school as Curcio's class, but it was taught by a different professor who did not provide any individualized feedback to students. Both instructors then blindly graded all the exams in the two classes. On average, students in Curcio's class, who received the individualized feedback, outperformed students in the control class that did not receive individualized feedback, even though there was no statistical difference between the two classes' average LSAT and undergraduate GPA. Interestingly, the benefit was principally concentrated among students with above-median LSAT scores or undergraduate GPAs.

As Curcio and her colleagues acknowledge, the 2008 study suffers from a number of important methodological limitations. Most significantly, the treatment of the control group in the study—the section that did not receive individualized feedback—differed not just in its receipt of individualized feedback, but also with respect to the instructor of the underlying material.

30. Greg Sergienko, *New Modes of Assessment*, 38 SAN DIEGO L. REV. 463, 485-86 (2001).
31. See Ramy, *supra* note 6, at 857-63; Sandra L. Simpson, *Riding the Carousel: Making Assessment a Learning Loop Through the Continuous Use of Grading Rubrics*, 6 CANADIAN LEGAL EDUC. ANN. REV. 35 (2011); Jay M. Feinman, *Teaching Assistants*, 41 J. LEGAL EDUC. 269, 272, 276 (1991).
32. Andrea A. Curcio, Gregory Todd Jones & Tanya M. Washington, *Does Practice Make Perfect? An Empirical Examination of the Impact of Practice Essays on Essay Exam Performance*, 35 FLA. ST. U. L. REV. 271 (2008) [hereinafter *Does Practice Make Perfect?*]. Preliminary results of the study and study design issues were discussed in an earlier article, Andrea A. Curcio, Gregory Todd Jones & Tanya M. Washington, *Developing an Empirical Model to Test Whether Required Writing Exercises or Other Changes in Large-Section Law Class Teaching Methodologies Result in Improved Exam Performance*, 57 J. LEGAL EDUC. 195 (2007).

Thus, one reasonable interpretation of the evidence is that Curcio was simply a better instructor than the instructor of the control-group class. Related problems arise from the fact that the two instructors identified the issues on which the ultimate exam would focus before teaching the underlying material in their classes. Curcio could therefore have unconsciously focused more attention on this subject matter in her teaching than the instructor of the control-group class. Another limitation of the study design is that it is impossible to know whether the instructors' grading of the exams was itself influenced by the interventions they gave and their involvement in the study. Although both instructors blindly graded all the exams, it is possible that they unconsciously provided outsized credit to exams that seemed to follow specific forms of guidance of the type contained in the individualized feedback provided to the treatment class. Alternatively, the instructors could have unconsciously used a broad grading curve that increased the gap between weak and strong students' grades to magnify the measured effect of individualized feedback.³³

Curcio's 2012 study, conducted with a different co-author, attempted to address some of these problems with the 2008 study design. It analyzed how providing a graded midterm and five ungraded short-answer exams affected the performance of Curcio's fifty-five-person evidence law class, which she taught in 2009.³⁴ Unlike the earlier study, the control group was a class that was also taught by Curcio—her 2008 evidence law class, in which she did not provide any enhanced feedback before the final exam. Both classes were given final exams consisting of two essays and either eighteen (in 2008, the control group) or fifteen (in 2009, the treatment group) short-answer questions. Eleven of the short-answer questions in the two exams were identical, and the study focused on students' performance on these questions across the two years. Students in the 2009 class that received the feedback outperformed students in the 2008 class by a statistically significant margin, even after controlling for undergraduate GPA and LSAT. To ensure consistency of grading across the two years, Curcio used a grading rubric. After blindly regrading fifteen exams from each year using the rubric, 295 out of 330 of the regraded questions received the same score as that assigned in the actual grading process.

Although the 2012 study eliminates the variability in instructors across control and treatment classes that complicated the 2008 study, it is subject to its own methodological problems. Perhaps most notably, Curcio taught the control and treatment classes in different years, meaning that the evidence

33. Curcio and her colleagues plausibly respond by noting that many of these potential concerns would be expected to improve all students' grades in the treatment group. Instead, they found that feedback primarily benefited the above-the-median LSAT score students. See *Does Practice Make Perfect?*, *supra* note 32, at 301.
34. Carol Springer Sargent & Andrea A. Curcio, *Empirical Evidence That Formative Assessments Improve Final Exams*, 61 J. LEGAL EDUC. 379 (2012). For a recent overview and synthesis of both studies, see Andrea A. Curcio, Gregory T. Jones & Tanya M. Washington, *Essay Question Formative Assessments in Large Section Courses: Two Studies Illustrating Easy and Effective Use*, in *EXPLORING LEARNING & TEACHING IN HIGHER EDUCATION* 349 (Mang Li & Yong Zhao eds., 2015) (also available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2617011).

could reflect the fact that Curcio simply did a better job teaching in 2009 than 2008 for reasons unrelated to feedback. Alternatively, Curcio may have more comprehensively covered the subject matter tested in the eleven common exam questions in the 2009 treatment class, a risk that is amplified by the fact that the eleven exam questions on which the study focuses had been written before the 2009 class. Yet another concern is that the 2009 treatment class may have had more time to devote to the eleven common questions than the 2008 control class because the nonoverlapping portions of the exam were easier or less time-consuming in 2009 than 2008. In addition to these control group issues, it is hard to eliminate the possibility that some or all of the measured effect might have been attributable to inconsistent grading. Not only is it possible that Curcio unduly rewarded exams that replicated feedback, but it is also possible that she unconsciously graded the 2009 exams more generously than the 2008 exams. Although the consistent blind regrading of fifteen exams from each year mitigates this concern, it does not eliminate it, because Curcio could have been unconsciously aware of whether she was regrading a 2008 or 2009 exam.

Although Curcio and her co-authors' two studies are subject to important methodological limitations, they provide, by far, the most developed empirical evidence in the literature on the impact of feedback on law students' grades. However, other relevant studies do exist.³⁵ For instance, one study dating back to 1981 evaluated the impact of more frequent testing of first-year law students.³⁶ The study randomly assigned seventy-five first-year students in a torts class to one of three conditions. In the first, students were tested four times with multiple-choice exams throughout the semester, with each exam counting toward twenty-five percent of students' final grade. In the second, students were tested twice through the semester, with each exam counting toward fifty percent of the final grade. In the final treatment, students were tested only once, at the end of the semester. All exams consisted of multiple-choice exams. The study found that students in the two sections that had previously been tested outperformed students in the section that were tested only once, at the end of the semester.³⁷ Because of the small sample size, however, the statistical significance of the results was limited. More importantly, it is hard to know from the study design whether the students who performed better after taking multiple exams did so merely because they became familiar with the question

35. Another article contains an instructor's description of his instruction of two different first-year classes (civil procedure and constitutional law), one in which he provided students with frequent assessments and one in which he did not. Charles A. Rees, *The "Non-Assessment" Assessment Project*, 57 J. LEGAL EDUC. 521 (2007). The article reports that students in the two classes performed equally well in the high-feedback and low-feedback classes. However, it is hard to draw conclusions from the article given that the control and treatment groups involve two different classes on different subject matters that were evaluated on the basis of different exams. *Id.* at 523.

36. Gary A. Negin, *The Effects of Test Frequency in a First-Year Torts Course*, 31 J. LEGAL EDUC. 673 (1981-82).

37. *Id.* at 674-76.

types favored by the instructor, or because they understood the underlying material better.³⁸ Finally, the study made no attempt to control for underlying student characteristics.

Another empirical study of relevance—particularly for its methodological approach, which also leveraged the random assignment of students to 1L sections—demonstrated that small class sizes tend to eliminate a preexisting gender gap in which male law students outperform female law students in first-year grades.³⁹ The study focused on the random assignment of Stanford law students to small and large 1L sections. Whereas men tended to receive higher grades than women in large section classes, no such disparity was observed in small section classes. Although the focus of the paper was on class size rather than instructor-provided feedback, the study also examined the impact of additional educational reforms that simultaneously changed the grading system to honors/pass from a 4.0 GPA scale and created a small (eighteen-person, split into four- to five-person teams) federal litigation class. Students in that class were given a wide range of writing assignments and instructors provided them with substantial feedback about their performance on these assignments. These reforms, the study found, eliminated the gender gap in student performance across the curriculum. However, the relative impact of feedback, independent of reductions in class size and changes in the grading system, is impossible to disentangle from the study because these reforms were implemented simultaneously.

II. Natural Experiment at University of Minnesota Law School

As at many law schools, first-year law students at University of Minnesota are assigned to one of several sections when they matriculate to the law school. These sections consist of a cohort of between forty and fifty first-year law students seeking a J.D. degree who take all their required first-year doctrinal classes together.⁴⁰ These classes are taught by a single slate of professors, and consist of contracts, civil procedure, torts, and constitutional law in the first semester, and property and criminal law in the second semester.⁴¹ Between 2011 and 2013, first-year law students were split into five sections. Because of decreasing enrollment numbers, the 1Ls entering the law school in fall 2014 and fall 2015 were split into only four sections. All first-year, graded classes at the

38. To be sure, formative assessment may be valuable even if it merely allows students to perform better on the ultimate summative assessment in a particular class. But this is much less likely to be the case when this effect is attributable to the idiosyncratic assessment approach favored by a particular instructor. To take a simple example, students who learn through various “formative assessments” that an instructor tends to construct multiple-choice questions where the correct answer is “none of the above,” and on the basis of this knowledge perform better on a final summative assessment, have not thereby learned a generalizable or intrinsically valuable skill that can be translated to other settings.

39. Ho & Kelman, *supra* note 17, at 305-06.

40. Some first-year classes also include LL.M. students or transfer students.

41. In addition to these six core classes, students select one elective course in the spring, have legal writing in both fall and spring, and must take a class in law in practice in the spring.

University of Minnesota Law School are subject to a mandatory distribution requirement, which has historically required the mean score in each first-year class to be between a 3.00 (“B”) and 3.33 (“B+”), and which more recently has been tightened to require a class mean of between 3.20 and 3.33.⁴²

To promote diversity of opinions and experiences in individual sections and to ensure that the mandatory distribution requirement does not unfairly affect students in different sections, law school administrators have consciously constructed sections so they include a balanced cross section of the law school’s overall entering class. To accomplish this, administrators run a “class-balancing program” that considers four student factors in constructing individual sections: (a) age; (b) LSAT score; (c) gender; and (d) ethnicity. The program is designed to distribute students of each matriculated class into sections that have roughly the same composition as the broader incoming class with respect to these four criteria. After they have run the balancing program, law school administrators occasionally tweak the results to ensure that, for instance, spouses or twins are not placed in the same sections.

At the same time law school administrators consciously seek to balance the student population across sections, the law school’s associate dean consciously seeks to balance the individual professors assigned to teach each section’s six required doctrinal classes. Of course, this process is less mechanical than that applied to students. The associate dean generally attempts to ensure that each section has a relatively diverse set of professors as well as a comparable set of relatively popular professors. However, none of the associate deans in recent years has made any attempt to balance professors with respect to whether they provide individualized feedback to students in their classes.⁴³ Yet individual professors who teach within the first-year curriculum at the University of Minnesota have varying practices on whether and how they provide students with individualized feedback. Although it is theoretically possible that some professors alter their teaching practices based on their perceptions about their students, this is unlikely in our data set because the composition of individual sections is smooth across all relevant characteristics and the instructors at the law school generally have time-consistent practices on providing feedback.

The associate dean of the law school generally aims to ensure that all required first-year classes are taught in classes of single sections. However, a variety of factors occasionally render this goal infeasible. In such cases, the associate dean generally combines two individual “component sections” into a “double-section” class taught by a single professor. The instructors of double-section classes are not generally aware of which students are members of which

42. The updated mandatory distribution requirement came into effect in 2014. In addition to requiring that the mean in the class fall within this domain, the new rules also require that at least thirty-three percent and no more than thirty-nine percent of students shall receive “A” level grades (A+, A, and A-). No such distribution requirement applied before 2014.

43. As a result of this study, the associate dean at the University of Minnesota Law School will in the future take into account the extent to which professors provide individualized feedback in constructing sections.

sections, and they teach and grade the two component sections collectively. As in most law school classes, the instructors of these double-section classes blindly grade all exams.

Collectively, these practices at the University of Minnesota Law School create a natural experiment on the impact of individualized feedback on law school performance.⁴⁴ In particular, they allow us to test whether instructors who provide individualized feedback to students in the first year of law school improve those students' ability to perform well in law school more generally. This natural experiment offers the potential to substantially advance the existing literature on feedback in law school for a number of reasons. First, unlike all prior research, this study design includes a natural control group: sections that have not been provided with individualized feedback but are paired in a double-section class with sections that have been provided with such feedback. Second, because this study focuses on a natural experiment, none of the instructors involved either in providing individualized feedback to students⁴⁵ or—more importantly—in grading the double-section classes could have been subconsciously biased by the desire to produce results consistent with their prior beliefs for purposes of the study. Third, and perhaps most notably, unlike prior research, this natural experiment allows us to test whether individualized feedback provided in one class affects first-year law students' performance in other first-year classes. If so, this limits the force of criticisms that individualized feedback merely results in “teaching to the test.”

III. Data

A. Data on Mandatory 1L Classes

To identify the individual instructors who taught each of the six required doctrinal classes for first-year law students from fall 2011 to fall 2015, we acquired from law school administrators complete copies of class schedules during this time frame. In total, there were 114 different mandatory traditional first-year doctrinal classes taught at the law school. Sixteen of these classes were double sections, and thirty-five instructors taught one or more of these classes.

B. Data on Individualized Feedback in Mandatory 1L Classes

Having identified all thirty-five instructors of mandatory doctrinal first-year classes during our time frame, we conducted an online survey of these

44. See Ho & Kelman, *supra* note 17, at 295 (noting that assigning students into sections, as opposed to allowing students to select their classes, can create a robust natural experiment setting).

45. One possible exception is that Schwarcz did provide individualized feedback in two first-year classes, and in one instance this resulted in a split-feedback double section. However, Schwarcz had only vaguely conceived of this project at the time he provided this feedback. More importantly, it is unclear how any knowledge of this project might have affected the feedback he gave to students, which consisted of a grade and written comments on a midterm exam that counted for ten percent of students' final grade in the class.

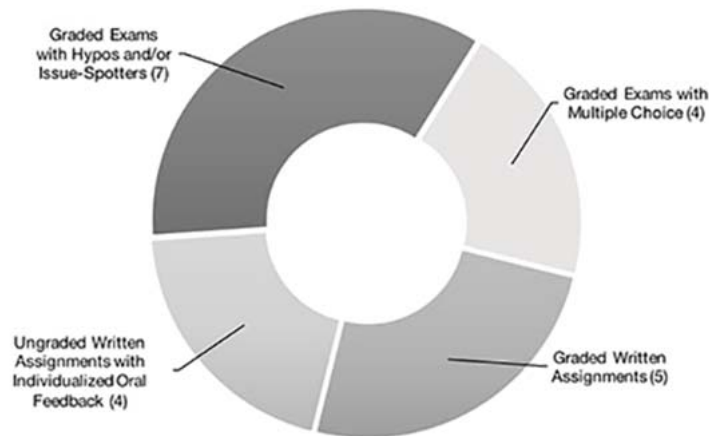
instructors. The survey asked a series of detailed questions about instructors' teaching practices and policies in each class they taught over this period. Thirty-four of the thirty-five instructors completed the survey.⁴⁶ Using these surveys, we identified all classes in which, before completion of the class, the instructor gave individualized feedback to students on exams⁴⁷ or assignments that contained the types of questions typically found on law school exams, such as issue-spotters or policy questions.⁴⁸ We excluded assignments that required students to produce "practical" documents, such as mock complaints or contracts, because these develop skills that are distinct from those tested on traditional law school exams.⁴⁹ We defined individualized feedback to include assigning grades to individual students' work products,⁵⁰ providing individualized written comments to students, and providing individualized or small-group oral feedback to students. By contrast, we did not consider individualized feedback to include generalized feedback in which instructors provided students with only a model answer, grading rubric, or in-class oral comments regarding strong answers or common mistakes.⁵¹ A substantial number of instructors in the first-year curriculum provided this type of generalized feedback on a mock exam or exam-like assignment. But only eight of the thirty-five instructors provided individualized feedback under these definitions in at least some of their classes, and twenty of the 114 different classes included such feedback.⁵²

The nature of the individualized feedback provided by the different instructors varied. Three of the eight professors in seven of the twenty classes

46. For the one nonresponding instructor, we were able to confirm that that instructor did not provide any individual feedback to students before the end of the semester by speaking with students who had been enrolled in the class.
47. An exam was defined in the survey as "generally subject to time restrictions and prohibitions on collaboration and [] designed to test knowledge, skills, or analytical abilities."
48. An assignment was defined in the survey as "not generally subject to time restrictions and [potentially] not . . . subject to prohibitions on collaboration." An assignment was deemed to be "exam-like" if it "require[d] students to write responses to issue-spotters and/or hypotheticals," or to "write responses to policy questions or similar essay-style questions."
49. Many professors at the law school required students to complete assignments involving these practical documents.
50. We considered an exam or exam-like assignment to be graded if "students received a grade reflective of their performance," even if those grades did not ultimately affect the final grade in the class. By contrast, a check mark or similar notation awarded to a student solely on the basis of a good-faith effort did not count as graded. We did not consider an exam to be graded if students "know in advance that their performance on the exam will not be graded."
51. Many instructors administered ungraded exams or assignments that were then discussed in class. Many instructors also provided students with model answers. We did not treat these classes as providing individualized feedback, and reserve for future study whether these techniques improved student performance.
52. Of the eight professors who gave individualized feedback, three taught four classes each, three taught two classes each, and two each taught one class.

administered graded exams to students containing issue-spotter or hypothetical questions, which were graded by the instructor and counted toward students' final grade in the class. Two of the instructors in four of the classes administered a multiple-choice exam that counted toward students' ultimate grades and was designed to test both basic knowledge (such as remembering rules and case holdings) and higher-order skills (such as applying rules to facts or analyzing hypotheticals). Two of the instructors in five of the classes gave students assignments that (a) were not subject to time restrictions and prohibitions on collaboration; (b) included traditional law school exam questions such as issue-spotters or hypotheticals; (c) were individually graded and commented on by the instructor; and (d) counted toward students' final grade in the course. Finally, one instructor who taught four classes gave students ungraded assignments consisting of traditional law school exam questions such as issue-spotters or hypotheticals and met with students in small groups or individually to discuss the answers. These results are summarized graphically below.

Figure 1: Types of Individualized Feedback Provided in First-Year Classes



C. Data on Student Grades and Characteristics

To assess the impact of individualized feedback on student performance, we acquired data from the law school registrar on all first-year J.D. students⁵³ at the University of Minnesota Law School from 2011 to 2015.⁵⁴ The data included each student's LSAT score, undergraduate GPA, date of birth, race/

53. We excluded from our data set all students seeking a degree other than a J.D., including an LL.M. We also excluded from our data set all transfer students enrolled in these classes. In very rare cases, students switched sections for idiosyncratic reasons. We dropped these students from our data set.
54. Before collecting this data, we received confirmation from the University of Minnesota's Institutional Review Board Human Subjects Committee that the data collection and study were exempt from review under federal guidelines 45 C.F.R. § 46.101(b) (2009), because it focused on instructional strategies in educational settings.

ethnicity, sex, country of birth, and law school section assignment. We also acquired from the registrar students' grades in all sixteen double-section 1L classes taught from 2011 to 2015. Although both data sets were anonymized, we linked them by using unique student identifiers that were derived from students' university ID numbers.

D. Data on Instructor Clarity

To assess the overall clarity of instructors' teaching techniques—a characteristic that we use as a control variable later in the study—we collected data on student end-of-semester teaching evaluations of faculty. These student evaluations of professors are generally completed by students in the last week or two of a class, and are not made available to instructors until final grades are turned in to the registrar. Throughout the examined period, one question asked students to rate their instructor on a six-point scale ranging from “strongly disagree” to “strongly agree” with regard to the following statement: “The professor presented the subject matter clearly.” The average score for all mandatory 1L classes during the period was 5.038. A “5” corresponded to “agree” and a “6” corresponded to “strongly agree.”

IV. Impact of Individualized Feedback on Law Student Performance

A. Identifying “Split-Feedback Double Sections”

Among the sixteen double-section classes taught during the studied time frame, we identified those in which, at the time of the final exam, one of the two component sections had received individualized feedback in at least one course and the other had not.⁵⁵ We refer to such classes as “split-feedback double sections.”⁵⁶ Among the sixteen double-section classes, eight met this description: three from fall semester and five from spring semester. In two

55. In fifteen of the sixteen double-section classes, students were given a final exam. The one double section in which this was not the case was not a split-feedback double section.

56. In the case of fall semester double sections, split-feedback double sections occurred only when an individual section in the double-section class had received individualized feedback in one of its single-section fall semester classes. For instance, sections A and B could be paired together in a fall-semester-double section constitutional law class, but Section A might have received individualized feedback in its concurrent fall-semester contract law class whereas Section B did not receive any such feedback. In the case of spring-semester double sections, a split-feedback double section could occur if one of the sections had received individualized feedback either in a prior fall-semester class or in one of its other spring-semester classes. Thus, if sections C and D were paired together in a spring double-section criminal law class, but Section C had received individualized feedback in its fall-semester contract law class while Section D had not received any within-semester individualized feedback in a prior or contemporaneous course, we counted this as a split-feedback double section. Our reasoning was that the literature suggests that feedback is most likely to promote learning when it is provided promptly, at a time when it can make a difference to a student's grade in a class. Students' receipt of their grades from the fall semester amid the spring semester does not satisfy this standard, such that we do not view second-semester 1L students to have received substantial formative feedback on their performance in law school based on their grades from fall semester.

of the eight of these split-feedback double-section classes, one component section received individualized feedback in two classes, and the other component section did not receive any individualized feedback. In the six remaining split-feedback double-section classes, one component section received individualized feedback in one class and the other component section did not receive any individualized feedback. This breakdown is illustrated in Table 1, below.

Table 1: Number of Feedback Classes in Component Sections

	Term	Component Section	Fall Feedback Classes	Spring Feedback Classes	Total Feedback Classes
Double Section A	Spring	A ₁ A ₂	1 0	0 0	1 0
Double Section B	Fall	B ₁ B ₂	0 1	— —	0 1
Double Section C	Spring	C ₁ C ₂	1 0	0 0	1 0
Double Section D	Spring	D ₁ D ₂	1 0	0 0	1 0
Double Section E	Fall	E ₁ E ₂	0 1	— —	0 1
Double Section F	Spring	F ₁ F ₂	0 1	0 1	0 2
Double Section G	Spring	G ₁ G ₂	1 0	0 0	1 0
Double Section H	Fall	H ₁ H ₂	2 0	— —	2 0

Although we treated classes with substantive multiple-choice exams as providing students with individualized feedback, all eight of the feedback component sections in fact received individualized feedback from their instructors on written exams or assignments.⁵⁷

57. In the case of the two component sections that received two classes of individualized feedback (double sections F and H), this feedback consisted of (a) a graded exam with issue-spotter and a multiple-choice exam; and (b) a graded exam with issue-spotter and ungraded written assignment with oral feedback. Among the six component sections that received one class of individualized feedback, this consisted of individualized feedback on graded exams with issue-spotters in four instances, and oral feedback on ungraded written assignments in two instances.

In total, 541 unique students received grades in these eight split-feedback double-section classes.⁵⁸ The characteristics of the students in our eight-class analysis are nearly identical to those of the law school's overall student population during the period reviewed (2011 through 2015). The groups have the same mean LSAT scores and undergraduate GPAs, and the differences in age (.5 years), percentage of male students (1.5), and percentage of white students (.2) are not statistically significant. This suggests that findings drawn from the double-section data are generalizable to the law school's overall student population.

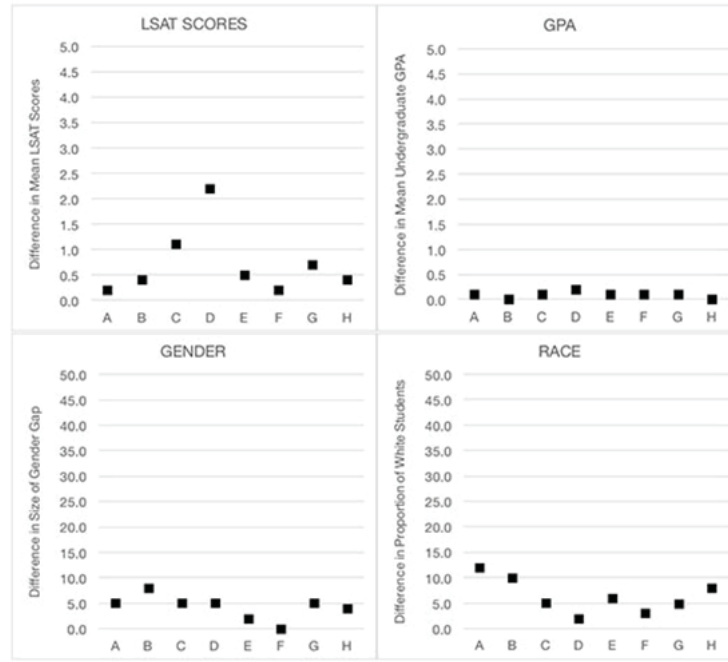
B. Confirming Similarities Among Component Sections of Split-Feedback Double Sections

A vital component of our natural experiment is the assumption that individual sections at the law school are roughly similar with respect to all relevant characteristics that they possess upon entering law school, such as LSAT and undergraduate GPA. For instance, if one component section of a split-feedback double section included a significantly greater number of high-LSAT scorers than the other component section, it would be harder to disentangle the cause of any differences in performance between these sections. By contrast, if we were to find that the component sections had roughly equivalent characteristics, then we can more confidently link differences in their performance in split-feedback double sections to their receipt of individualized feedback.

Although law school personnel specifically design sections to be similar in key attributes that could affect law school performance, we confirm here that each of the two component sections for our eight double sections do indeed possess similar key attributes. Figure 2 illustrates graphically the even distribution among component sections within each split-feedback double-section class with regard to LSAT scores, undergraduate GPAs, gender gap (i.e., the size of the percentage gap between male and female students), and race/ethnicity (i.e., the percentage of the students in each component section who are U.S.-born white). Significance testing finds no statistically significant differences among the students in any of the eight pairs of component sections.

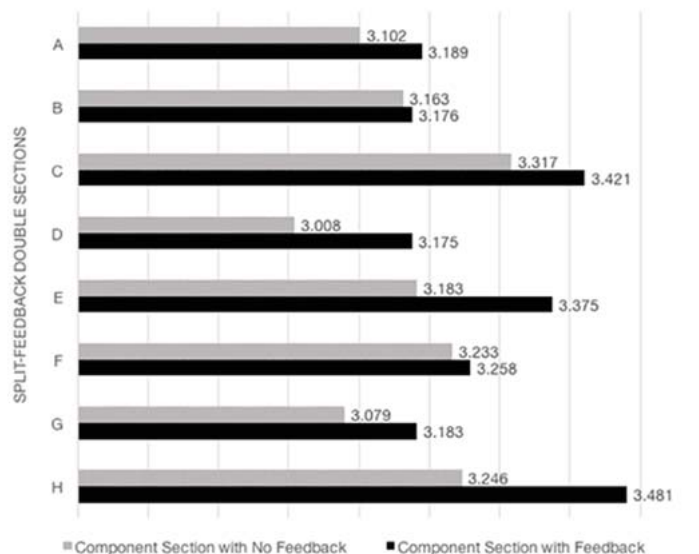
58. A total of 122 students took two double-section classes, meaning that the aggregate number of graded students throughout the two classes was 663, while the number of unique students was 541. In one instance, a single section was the feedback component section in two split-feedback double-section classes, though the no-feedback component section varied across these two split-feedback double-section classes. In another instance, a single section was the no-feedback component section in two split-feedback double-section classes, though the feedback component section varied across the two split-feedback double-section classes.

Figure 2: Differences between Students in Component Sections



C. Performance of Component Sections of Split-Feedback Double Sections

Having now identified the eight split-feedback double sections and confirmed that, in each case, the two component sections possess similar relevant attributes, we now examine the aggregate performance of the component sections. If individualized feedback in one law school class does not affect students' performance in other law school classes, we would expect that the two component sections would generally perform equally well in the split-feedback double sections. By contrast, if individualized feedback in one law school class positively affects students' performance in other law school classes, we would expect that the feedback component sections would generally outperform the no-feedback component sections. Figure 3 illustrates that in all eight split-feedback double sections, the feedback component section outperformed the no-feedback component section in the split-feedback double section.

Figure 3: Mean Grades for Component Sections

Although none of the differences in mean GPAs between a feedback and no-feedback component section in a single split-feedback double section is statistically significant on its own, the likelihood of the feedback component section randomly outperforming the no-feedback component section in all eight split-feedback double sections is $1/(2^8)$, or one out of 256. Moreover, when all eight split-feedback double sections are aggregated together, the difference in mean GPAs between the no-feedback group (3.165) and the feedback group (3.283) is statistically significant ($p = .010$).⁵⁹

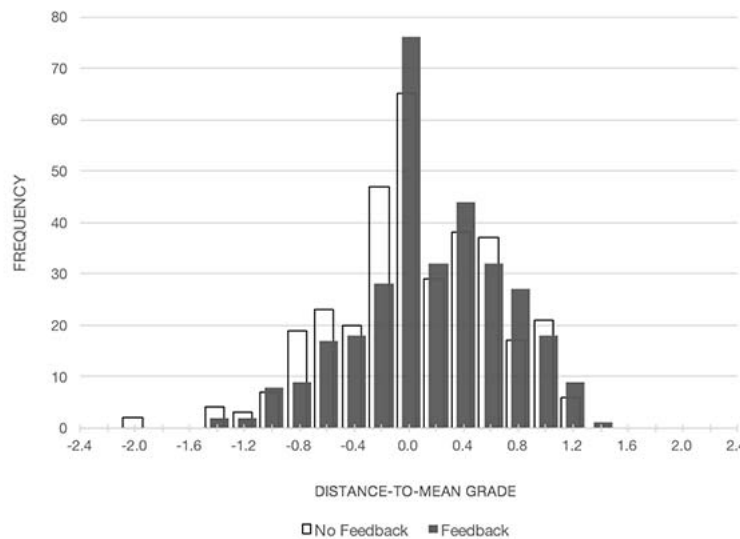
Spurred by these initial results showing a disparity in GPAs, we develop a new dependent variable, which is defined as the distance between each student's grade in a split-feedback double-section class and the mean grade for that class. This variable allows us to examine students' performance in all eight of the split-feedback double-section classes in the aggregate despite the variance in the mean grade in each instructor's class. Under this new metric, the students in the feedback component sections continue to outperform their no-feedback counterparts. Specifically, we find that the average distance-to-

59. After a preliminary version of this paper was made available online, Professor Michael Asimow contacted us and informed us that he had informally conducted a nearly identical experiment at UCLA Law School. In particular, he gave students in his first-semester contract law class a graded midterm with extensive feedback, which counted for about twenty percent of their final grade. His colleague taught a different section of contract law using the traditional format, with only a final exam. Both classes then fed into a common torts class, which was taught in the spring semester. Consistent with this study's findings, Professor Asimow's students performed much better in the common torts class than did the students who had the alternative contract law professor.

mean score for the students who were in a feedback component section is .0432, while the score for the no-feedback group is -.0453. In other words, those in the feedback group are, on average, above the mean in their respective classes, while those in the no-feedback group are below it. The mean difference between the groups of .0885 is statistically significant ($p = .049$).

These disparities in performance are reflected in the histogram in Figure 4, which graphs the frequency of students in the feedback and no-feedback component sections at various points along the distance-to-mean spectrum. The solid gray bars represent students in the feedback component sections; the unfilled bars represent students in the no-feedback component sections. A normal distribution of grades would be concentrated around the middle value (0.0), with even numbers of students on either side, growing progressively less frequent toward the extreme low and high values. Here, the distribution for the students in feedback component sections is notably heavier in the above-mean part of the graph, while the distribution for the students in the no-feedback component sections is considerably overloaded in the below-mean section.⁶⁰

Figure 4: Distribution of Grades for Feedback and No-Feedback Groups



Notably, disparity in performance between the feedback and no-feedback groups appears most pronounced at the bottom end of the spectrum, for the lowest-performing students. To examine this apparent difference more closely, we analyzed the extent to which students in a no-feedback component section were overrepresented in the group of students who scored .2 or more below the

60. Because the no-feedback group (N=339) was slightly larger than the feedback group (N=323), we generated a relative frequency histogram in order to ensure that the distribution patterns were replicated when adjusting for the different population sizes. We found no appreciable difference in the distribution patterns between the two histograms.

class mean in split-feedback double-section classes. We found that thirty-eight percent of the students in no-feedback component sections were in this group, while only twenty-seven percent of the students from the feedback component sections scored .2 or more below the mean. This difference in proportions is highly statistically significant ($p = .001$).

Because of the limited number of split-feedback double sections, our data do not allow us to confidently test the relative impact of different types of individualized feedback. As described above, our analysis lumped together a number of different types of individualized feedback, including oral and written feedback, feedback on midterm exams and assignments, feedback on issue-spotters and policy questions, feedback on graded and ungraded assignments, and feedback delivered through multiple-choice exams. Of course, we have some intuitions about which of these forms of feedback are most and least valuable. However, our results provide only limited evidence to back up these intuitions, with one caveat: As noted above, each of the feedback component sections happened to receive instructor-provided individualized feedback on their written work product in at least one class.⁶¹ For this reason, it is reasonable to interpret our results to apply predominantly to this form of individualized feedback, rather than to objective multiple-choice exams.

Although we cannot offer much evidence regarding the relative effectiveness of different forms of individualized feedback, this fact actually strengthens our results in an important sense. Assume, as is likely, that some types of individualized feedback—such as written comments on a graded midterm exam—are relatively more effective than others. If so, the implication is that providing this more effective type of feedback to students would have a much greater effect than what we measure here, precisely because we lump together both more and less effective forms of feedback.

Our data also do not allow us to disaggregate the impact of providing individualized feedback in the same semester as a split-feedback double section or in an earlier semester. As suggested in Table 1, in half of the split-feedback double sections, students received individualized feedback only in the prior semester, whereas in the other half students received feedback in the same semester as the split-feedback double section.⁶² It may well be that the impact of feedback differs in these two cases. Once again, we believe that this point actually strengthens our results, suggesting that if we could identify the optimal timing for delivering feedback and act accordingly, the impact on student performance would be greater than that we measure by lumping together feedback delivered at different times.

The limited number of split-feedback double-section classes also precludes us from confidently assessing the marginal impact of providing students with individualized feedback in more than one class. Recall that the feedback component sections in double sections H and F received individualized

61. See *supra* Part IV.A.

62. See *supra* Table 1.

feedback in two classes rather than one.⁶³ But the relative size of the GPA gap in these two double sections was mixed: The GPA gap among component sections was highest among all eight double sections in Double Section H, but it was second-lowest among the eight double-section classes for Double Section F. Also relevant is the relative performance of component sections in one double-section class in which one component section received individualized feedback in two classes and the other in only one class. We did not designate this double section as one of our eight split-feedback double sections because both component sections received individualized feedback.⁶⁴ However, the component section that received individualized feedback in two classes outperformed the component section that received only one class of individualized feedback, with the former having an aggregate GPA of 3.383 and the latter an aggregate GPA of 3.284. In sum, while some preliminary results indicate that a second dose of individualized feedback may improve student outcomes, a reliable assessment of this issue will require further study.

D. Regression Analysis of Performance of Component Sections in Split-Feedback Double Sections

Although the results so far provide compelling evidence that individualized feedback in a component section's class improves that section's performance in the split-feedback double section, we extend the analysis using a linear regression model. Our dependent variable is the distance-to-mean-grade score used previously for each student in a split-feedback double section. Our independent variables are: LSAT score; undergraduate GPA; gender (0=female, 1=male); two dummy variables for race/ethnicity (U.S.-born nonwhite student and non-U.S.-born student); and feedback (0=no-feedback component section, 1=feedback component section).⁶⁵ We choose these variables because research has suggested that each one does, or might, operate as a predictor of law school performance.⁶⁶ Table 2 presents these regression results.

63. See *supra* Table 1.

64. As described earlier, we designated a double-section class as a split-feedback double section only if one of the two component sections had received individualized feedback in at least one course and the other had not received any individualized feedback at all. *Supra* text accompanying note 54.

65. Nonwhite includes anyone who identifies as Asian, African-American, Hispanic, Native American, or two or more.

66. See William D. Henderson, *The LSAT, Law School Exams, and Meritocracy: The Surprising and Undertheorized Role of Test-Taking Speed*, 82 TEX. L. REV. 975, 981 (2004).

Table 2: Factors Influencing Student Grades in Split-Feedback Double-Section Classes

LSAT Score	.036 **	(.005)
Undergraduate GPA	.295 **	(.050)
Gender	-.008	(.047)
U.S.-born non-white	-.217 **	(.025)
Non-U.S.-born	-.242 *	(.069)
Feedback	.134 **	(.029)
Number of observations	562	
R ²	.221	

Note: Table represents results of a linear regression model with “distance to mean grade” as the dependent variable. Parentheses contain robust standard errors, clustered by double-section class.

+ $p < .10$; * $p < .05$; ** $p < .01$.

The regression provides strong support for our hypothesis that feedback matters. As shown in the table, receiving feedback is a statistically significant predictor of student performance in the anticipated (positive) direction. The model predicts that a student who receives feedback in a component section will see his or her grade in the double-section class improve by .134 points (so, for example, from a 3.000 to a 3.134) when all the other variables are held constant. This is the same impact on a student’s predicted grade as a 3.7-point increase in that student’s LSAT score. With regard to the other independent variables, LSAT and undergraduate GPA are, as one might expect, strongly significant predictors of grade performance. Interestingly, while gender is not statistically significant, race/ethnicity does appear to play a role in student grade performance. Specifically, U.S.-born nonwhite students and non-U.S.-born students are predicted to perform less well in split-feedback double-section classes when holding other factors constant. With respect to non-U.S.-born students, we suspect that this effect largely reflects the fact that English is not likely to be this group’s native language. Although it is obviously much more difficult to explain why U.S.-born minority students perform less well than their peers even after controlling for other factors, this trend has been documented at various other law schools by numerous other studies.⁶⁷

67. See, e.g., Alexia Brunet Marks & Scott A. Moss, *What Predicts Law Student Success? A Longitudinal Study Correlating Law Student Applicant Data and Law School Outcomes*, 13 J. EMPIRICAL STUD. 205, 243 (2016) (finding racial disparities in law school performance in a study of two law-different law schools from 2005 to 2012, even after “controlling, better than prior studies do, for not only academic ability on standardized tests (i.e., LSAT) and prior academic performance (i.e., UGPA), but also a number of other variables relevant to academic credentials, such as college quality, college major, and UGPA trajectory”); John Fordyce et al., *Predicting First-Year Law School Performance: The Influences of Race, Gender, and Undergraduate Major*, 43 EASTERN ECON. J. 64 (2017) (similar).

Not only is the effect of feedback statistically significant, but it is also meaningful in its magnitude. As noted above, students who receive individualized feedback in one class experience about a .134-point increase in their expected grade in a different class in which only half the student population has received individualized feedback. To appreciate the magnitude of this effect, recall that it is the result of receiving individualized feedback in one—or at most two—classes during the first year of law school. Moreover, the difference in GPA of .1 on a 4.0 scale can certainly impact student outcomes.⁶⁸ Perhaps most notably, it can affect class rank or reported class quartile. To take one example, in a typical year at the law school, a .134-point increase in GPA can improve a student's class rank by as much as fifty places, which in turn can mean the difference between being ranked in the second versus the first quartile. Class rank and percentile are often important for firms reviewing applications for coveted summer associate positions, and both weigh just as heavily in the highly competitive universe of judicial clerkships. In short, we are persuaded that the differences in mean grades among the component sections are, in terms of their real-world implication, very meaningful.⁶⁹

To further investigate whether the impact of feedback is felt equally across the student population or is instead concentrated among subpopulations of students, we conducted a second set of regressions using the same variables. But this time we divided the students into two groups based on their LSAT scores and undergraduate GPAs. To do so, we relied on the law school's formula for calculating a student's LSAT/GPA "index,"⁷⁰ which is designed to predict students' first-year grades at the law school using past students' LSAT, UGPA, and first-year grades. We divided the students into a below-median index group and an above-median index group, and then estimated the effects of our independent variables on their distance-to-mean grades in the double-section classes.⁷¹

68. Ho and Kelman develop this point extensively, showing (among other things) that small differences in GPAs at Stanford Law School have meaningful impacts on students' chances of receiving a judicial clerkship. *See* Ho & Kelman, *supra* note 17, at 300.

69. Of course, we acknowledge that providing individualized feedback to all students—as we suggest above and below—would not unambiguously increase the employment or clerkship prospects of all students, as it would in part simply shift the curve upward. The point here is simply that the impact of providing individualized feedback to only some students is to meaningfully improve their performance in their other classes, which is itself reflected in the different employment and clerkship prospects that flow from the improved performance in those classes.

70. The law school's LSAT/GPA index formula is: $(\text{LSAT score} \times 0.035) + (\text{GPA} \times 0.368) - 3.688$.

71. Recall that prior research had found that feedback had a stronger effect on students who scored at or above the median LSAT score. *See Does Practice Make Perfect?*, *supra* note 32, at 300–01. By contrast, some general literature on formative feedback suggests that it can sometimes have a more positive impact on relatively low-performing students. Paul Black & Dylan Wiliam, *Inside the Black Box: Raising Standards Through Classroom Assessment*, 80 PHI DELTA KAPPAN

As shown in Table 3, the effect of feedback is larger in magnitude and enjoys greater statistical significance for the group of students who were below-median performers in terms of LSAT and undergraduate GPA.⁷² These results reinforce the conclusion, suggested earlier, that feedback appears to be having a stronger effect among comparatively low-performing students at the University of Minnesota Law School.⁷³

Table 3: Factors Influencing Student Grades, by LSAT/GPA Index Scores

	Below-Median LSAT/GPA Index		Above-Median LSAT/GPA Index	
LSAT score	.041 **	(.013)	.038 **	(.009)
Undergraduate GPA	.290 +	(.146)	.502 **	(.104)
Gender	-.004	(.072)	.007	(.064)
U.S.-born non-white	-.290 **	(.068)	-.096	(.095)
Non-U.S.-born	-.322 **	(.052)	-.141	(.119)
Feedback	.172 *	(.053)	.082 +	(.041)
Number of observations	278		280	
R ²	.198		.097	

Note: Table represents results of linear regression models with “distance to mean grade” as the dependent variable. Parentheses contain robust standard errors, clustered by double-section class.

+ $p < .10$; * $p < .05$; ** $p < .01$.

E. Variance in Component Sections’ Performance in Split-Feedback Double-Section Classes and Non-Split-Feedback Double-Section Classes

To this point, the analysis has focused exclusively on the relative performance of component sections in split-feedback double-section classes. This focus is sensible, as these are the only double-section classes in which a

^{139, 141} (1998) (reviewing approximately 580 sources related to formative assessment and finding that “[m]any of these studies arrive at another important conclusion: that improved formative assessment helps low achievers more than other students.”).

72. We also note an additional interesting pair of results from these regressions. The LSAT score of students who are below the University of Minnesota median LSAT/GPA index is a statistically significant predictor of grade performance at the one percent level, while their undergraduate GPA is statistically significant only at the 10 percent level. Recall that for the combined group, both LSAT and GPA were statistically significant coefficients at the one percent level.

73. This result appears to be in some tension with Curcio’s finding, described earlier, that feedback had a stronger effect on students who scored at or above the median LSAT score. Importantly, though, the median LSAT score at University of Minnesota Law School during the relevant period differed from the median LSAT score at Georgia State Law School during Curcio’s study, which was 161 in 2010. This suggests the possibility that feedback might be most beneficial for students within a specific range of LSAT scores.

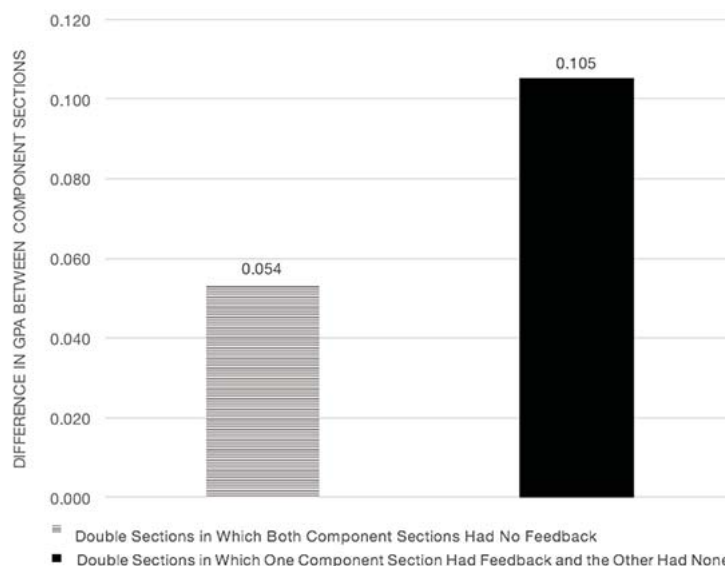
natural experiment occurs by virtue of one and only one of the two component sections having received feedback by the time of the final exam. At the same time, to the extent that individualized feedback is having the effect we identify, one might reasonably hypothesize that the gap in performance between component sections' performance in split-feedback double-section classes—where one and only one component section received individualized feedback—would be larger than the gap in performance between component sections where there was no split in component sections' receipt of individualized feedback.

To determine whether this is indeed the case, we calculate the mean difference in GPAs among component sections in both split-feedback double-section classes and non-split-feedback double-section classes. Recall that there were a total of sixteen double-section classes in our data set, and eight of these classes were split-feedback double-section classes. Among the remaining eight non-split-feedback double-section classes, we focus the comparison on the four non-split-feedback double-section classes in which neither component section had received any individualized feedback at the time of the exam. We do this because we have no reason to believe that all forms of individualized feedback are equally effective, and thus we do not necessarily expect the baseline divergence of component sections' GPAs in double-section classes to be reflected when both component sections received individualized feedback.

As illustrated in Figure 5, the gap in GPAs among component sections is about twice as large in split-feedback double section as non-split-feedback double sections. This once again suggests that individualized feedback is improving student performance in split-feedback double-section classes. Although the analysis in Figure 5 focuses on the four non-split-feedback double-section classes in which neither component section received any individualized feedback, we note that the result is virtually identical when we include the three additional non-split-feedback double-section classes in which both component sections received individualized feedback in the same number of classes.⁷⁴

74. In either scenario, it is appropriate to exclude the one double-section class, described in Part IV.C, in which one component section received individualized feedback in one class and the other component section received individualized feedback in two classes. Recall that the component section that received feedback in two classes had a mean GPA of 3.383 in the double-section class, whereas the section having only one class of individualized feedback had an aggregate GPA of 3.284. This difference of .099 is more consistent with that found in split-feedback double-section classes, which we take to buttress our findings rather than to undermine them.

Figure 5: GPA Differences between Component Sections in Split-Feedback and No-Feedback Double Sections



V. Addressing Objections

The most important objections to our analysis involve the prospect that instructor feedback, as we have defined it, is not the critical factor influencing outcomes, but instead simply correlates with the key driver of these outcomes. We tackle the two most important versions of this argument below.⁷⁵

A. Correlation of Professors Who Provide Individualized Feedback and Professors Who Are Effective for Other Reasons

Perhaps the most important objection to our results is that professors who provide individualized feedback also tend to possess some additional characteristic that is itself leading students in feedback component sections to outperform those in no-feedback component sections. Most intuitively, professors who provide individualized feedback may simply be “better” professors than those who do not provide such feedback, and it may be this

75. Of course, there are plenty of additional objections or complications beyond those that we specifically address in the text. For instance, while all students must take legal writing in both semesters and a legal-practice course in the spring, we do not control for teacher quality in those courses in measuring students’ performance in split double-section classes. (The content of legal writing and legal practice classes as well as the assignments required of students in these classes is standardized for all 1L students.) However, we view this limitation as minor, because there is no reason to suspect that any differences that might exist across students’ legal writing or legal practice classes would correspond to the differences in individualized feedback that sections received in their doctrinal classes. As such, it is hard to see how differences in legal writing or legal practice classes could produce the results that we observe.

fact—rather than the provision of individualized feedback—that leads their students to perform better in their other law school classes.

Before empirically evaluating this objection, we note that its intuitive force is limited because each section has a slate of either four (for fall semester) or six (for spring semester) different professors, and the associate dean consciously attempts to balance the quality of professors across sections. Thus, even if it were the case that professors who provide individualized feedback are disproportionately better professors than those who do not, it would not follow that component sections that received individualized feedback would tend to have a higher-quality slate of professors than component sections that did not receive individualized feedback. Instead, it would follow only that one of the high-quality professors in feedback sections also gives students individualized feedback, whereas none of the high-quality professors in the no-feedback sections provides students with individualized feedback.

To empirically study this issue more carefully, we collected data from student evaluations on the extent to which instructors presented material clearly. The clarity of a professor's presentation, we reasoned, reflects that professor's overall teaching skills, but should not be strongly related to whether the professor also provides individualized feedback to students. Thus, a professor might present material very clearly while providing no individualized feedback, or a professor might present material unclearly but provide individualized feedback. Moreover, a number of studies have found that student evaluations of teachers are "moderately correlated with independent measures of student learning and achievement."⁷⁶ At the same time, of course, ample research demonstrates biases in student evaluations of teachers, which include biases against female instructors and instructors who give lower grades.⁷⁷ However, research suggests that students do not tend to penalize professors who give students more time-consuming or difficult assignments, as long as students view these assignments to be appropriate for the course.⁷⁸ To be sure, it is still possible that the student evaluations of feedback professors might be biased downward for other reasons. Perhaps students who receive negative feedback on their work are disproportionately likely to "penalize" instructors with negative teaching evaluations. But we find it equally plausible that student evaluations of clarity will be biased upward for feedback professors in response to student appreciation of feedback. We therefore do not expect

76. BARBARA GROSS DAVIS, *TOOLS FOR TEACHING* 534 (2d ed. 2009). See also IDEA PAPER NO. 50, *STUDENT RATINGS OF TEACHING: A SUMMARY OF RESEARCH AND LITERATURE* 5 (2012).

77. See, e.g., Anthony C. Krautmann & William Sander, *Grades and Student Evaluations of Teachers*, 18 *ECON. EDUC. REV.* 59 (1999) (showing that higher-education instructors who give better grades get better teacher evaluations); Susan A. Basow & Nancy T. Silber, *Student Evaluations of College Professors: Are Female and Male Professors Rated Differently?*, 79 *J. EDUC. PSYCHOL.* 308 (1987) (finding that female instructors in higher-education settings received systemically lower teaching evaluations than male instructors).

78. See John A. Centra, *Will Teachers Receive Higher Student Evaluations by Giving Higher Grades and Less Course Work?*, 44 *RES. HIGHER EDUC.* 495, 515 (2003).

that biases in student teaching evaluations will tend to systematically affect feedback professors relative to no-feedback professors.

Using these data, we then analyzed whether discrepancies in the number of high-clarity professors assigned to each component section affected the performance of these component sections in split-feedback double sections. To accomplish this, we developed a new “clarity” variable to measure how many high-clarity instructors each component section had been assigned before or concurrently with the split-feedback double section. The variable reflected the percentage of instructors for each class taught to a component section (other than the instructor of the split-feedback double section) who received better-than-average clarity scores on end-of-semester student evaluations. We reasoned that if instructor feedback were simply proxying for instructor clarity, the feedback variable would become insignificant, and the clarity variable would become significant in the new regression analysis.

**Table 4: Factors Influencing Student Grades
(Clarity Variable Included)**

LSAT Score	.036 **	(.005)
Undergraduate GPA	.294 **	(.051)
Gender	-.009	(.047)
U.S.-born non-white	-.216 **	(.025)
Non-U.S.-born	-.242 *	(.069)
Feedback	.131 **	(.030)
Clarity of Instructor	.054	(.061)
Number of observations	562	
R ²	.221	

Note: Table represents results of a linear regression model with “distance to mean grade” as the dependent variable. Parentheses contain robust standard errors, clustered by double-section class.

+ $p < .10$; * $p < .05$; ** $p < .01$.

As Table 4 reports, including the instructor clarity variable in the regression analysis did not substantially alter the statistical significance of the feedback variable. Moreover, the instructor clarity variable was itself not statistically significant. Coupled with the associate dean’s regular attempts to balance the quality of professors across first-year sections, we believe these results provide reasonably strong evidence that feedback component sections outperform no-feedback component sections for reasons unrelated to no-feedback-related elements of instructor quality.

B. Correlation of Feedback Component Sections and Limited End-of-Semester Exams

Another objection stems from the possibility that instructors who provide individualized feedback during the semester have less burdensome exams at

the end of the semester. If so, then it may be that the reason students enrolled in feedback component sections perform better in split-feedback double-section classes than the students enrolled in no-feedback component sections is that they have fewer competing demands placed on them during finals period. For instance, if Section A has a high-feedback instructor who does not give a final exam and Section B has only traditional law school classes in which there is a single final exam, Section A may outperform Section B in the combined split-feedback double section simply because it has one less exam to study for.

To examine this possibility, we identified whether any of the individualized feedback professors forgo a final exam in their class. This was indeed the case for one high-feedback instructor who teaches a fall-semester class that does not include a final exam, but does include a final paper. Moreover, this single instructor provided the feedback for the feedback component section in four of our eight split-feedback double-section classes.

On further investigation, however, virtually no plausible argument exists that the strong performance of this instructor's students in split-feedback double-section classes was attributable to his lack of a final exam. Most importantly, this is because in three of the four instances in which this instructor taught one of the component sections of a split-feedback double section, the split-feedback double section was a spring class. The lack of a final exam in a fall-semester class could not plausibly affect the amount of study time students have for their spring-semester finals.

The issue is more complicated with respect to the single instance in which this instructor taught a component section of a split-feedback double section, in which the split-feedback double section was itself a fall-semester class. In this instance, the students in the feedback component section did indeed have one less exam to study for than their counterparts in the no-feedback section. To examine whether this biased our results, we identified instances in which the instructor in question taught a component section of a fall double-section class that was not itself a split-feedback double section because both component sections had benefited from a class providing individualized feedback. For instance, the instructor at issue might have taught Section A contract law in the fall, meaning that Section A did not have a contracts final exam. Meanwhile, Section A might have been paired with Section B in a double section of tort law, but the tort law class would not be counted as a split-feedback double section because Section B happened to have a fall civil procedure instructor who provided individualized feedback. To the extent that the instructor's students outperformed their counterparts in a double-section class because of their reduced demands during final period, we would expect that this effect would exist even in double sections such as this, where both sections received feedback. By contrast, to the extent that feedback was driving the effect rather than the reduced end-of-semester load on students, we would expect that double sections involving this professor in which both component sections received individualized feedback would not exhibit a split in the performance of the two component sections.

Two of the sixteen double sections have the basic characteristics described above: They were taught in the fall, one of the component sections was taught by the individualized feedback professor who does not give a final exam, but the double was not a split-feedback double section because both component sections received individualized feedback in their other classes. In the first of these special double sections, the mean GPAs were 3.33 for the component section receiving one less exam in the fall, and 3.31 for the other component section, for a total difference in GPA of .02. In the second of these double sections, the mean GPAs were 3.29 for the section receiving one less exam and 3.26 for the other component section, for a total difference in mean GPAs of .03. Although these differences suggest that it is possible that students who were required to study for one less exam benefited marginally in their overall performance, the magnitude of this effect seems to be no more than twenty-five percent of the magnitude of the average feedback effect we find above. These results suggest that in the single split-feedback double-section class in which the feedback component section had one less exam than the no-feedback component section, the predominant explanation resulted from the discrepancy in feedback rather than the discrepancy in the number of exams for which these two sections were required to study. In any event, given that this issue affects only one of our eight split-feedback double-section classes, we are confident that it does not undermine our results.

VI. Normative Implications

This article's findings have a number of important implications for the structure of legal education. First, they suggest that providing individualized feedback to law students may well promote their acquisition of the legal skills that traditional law school exams are designed to test. However, further research will be required to assess whether and when this outcome obtains. Our results do clearly demonstrate that providing students with individualized feedback in a core doctrinal class improves their ability to produce high-quality law school exam answers in general. Moreover, because students perform better in law school classes different from those in which they receive feedback, our results cannot be dismissed as simply documenting the impact of "teaching to the test" in a particular class. The most straightforward interpretation of these results thus appears to be that individualized feedback promotes students' acquisition of the skills that professors generally intend to teach and test, such as communicating clearly in writing, recognizing important legal issues, synthesizing applicable legal precedent, developing persuasive policy arguments, and marshaling the most relevant facts to support any conclusions. Indeed, this is consistent with anecdotal evidence on the content of the individualized feedback that professors in our study provide to students, which often focuses on these types of issues.

However, our results are open to at least two alternative interpretations that suggest that individualized feedback may not, in fact, be directly promoting students' acquisition of legal skills. One possibility is that the

primary mechanism by which feedback improves students' performance is by encouraging students who are not working hard enough or who are studying inefficiently to alter their approach. If so, individualized feedback may improve only students' relevant skill set indirectly, and alternatives, such as teaching students how to approach law school studying more systematically, may be more efficient or effective. A second, more troubling possibility is that individualized feedback may merely improve students' ability to "game" law school exams for reasons that have nothing to do with their development of relevant legal skills. The plausibility of this interpretation depends in large part on how well law school exams test skills that really matter to practicing lawyers. If traditional law school exams do a very good job on this front, it follows that students will generally be unable to use feedback to adopt approaches that improve their exam-taking abilities but are unrelated to their acquisition of relevant legal skills. By contrast, students may be able to use feedback in precisely this way to the extent that law school exams test skills or knowledge not relevant to the practice of law. Although we believe traditional law school exams do a relatively good job of testing a number of important legal skills for prospective lawyers, we acknowledge that this depends on the quality of individual exams as well as myriad other factors.⁷⁹

Second, this article's results suggest that instructors' provision of individualized feedback to students can have important distributional consequences. Because individualized feedback appears to disproportionately benefit students who would otherwise be at the lower end of the law school grade distribution, such feedback is likely to improve the job prospects of the law students most at risk of facing bleak employment outcomes. To be sure, the presence of a mandatory curve in most law schools means that improving the performance of these students will also affect the grades of other students. But this may ultimately be a socially beneficial trade-off, given the large disparities in job outcomes among high-performing and low-performing students at many law schools. Further reinforcing this point is that low-performing law students are also disproportionately likely to be saddled with large amounts of law school debt. This is because virtually all law schools give substantial tuition assistance to students with strong credentials while charging "sticker price" only to students who enter law school with the lowest credentials. Given the comparatively large debts and relatively limited job prospects facing many students who perform poorly in law school, law schools have a special obligation to improve the performance of this subset of students.

Third, our results suggest that individualized feedback can also have important fairness implications by artificially improving the performance

79. Some commentators have suggested that the only purpose of law school exams is simply to "sort" students so that employers can make better hiring decisions. However, we believe this cynical characterization of law school exams to have limited merit. Cf. Jeffrey Evans Stake et al., *Income and Career Satisfaction in the Legal Profession: Survey Data from Indiana Law School Graduates*, 4 J. EMPIRICAL LEG. STUD. 939, 970, 973 (2007) (finding that five years after law school, "each additional 0.1 on the graduate's GPA yields \$3,449 in additional annual income," but by fifteen years after, LGPA has no effect on income).

of some law students over others. As has traditionally been the case at the University of Minnesota Law School, most law schools do not currently make any effort to ensure that all students receive roughly equivalent levels of individualized feedback. As a result, law students assigned to professors who happen to provide such feedback have an advantage in their law school classes over law students not assigned to such professors. This system undermines the meritocratic nature of law school. It also undermines the reliability of law school grades as a signal to prospective employers on students' legal skills.

Fourth, and more speculatively, it is possible that providing law students with more individualized feedback would help improve law students' experience during law school. In particular, such feedback could enhance the ability of students to study effectively and reduce the sense of frustration that some students at the bottom of their class feel throughout law school. On the other hand, it is possible that enhanced feedback could have the opposite impact, by increasing students' workloads. Indeed, at least one survey found that law students at the start of law school indicate a preference for multiple graded and ungraded assignments, but have less enthusiasm for that approach toward the end of their first year of law school.⁸⁰ It is also possible that earlier feedback may increase law students' stress by creating "winners" and "losers" among law students before they receive their first-semester grades.

We believe that, when considered against the preexisting empirical research on the benefits of individualized feedback in higher education generally, and in law school in particular, our results suggest that law schools should systematically provide first-year law students with individualized feedback in at least one "core" doctrinal first-year class before final exams. This limited intervention improved students' performance in their other classes in the current study. Moreover, the costs of this intervention are minimal. For some law schools, this reform would simply require more thoughtful assignment of professors to individual sections to more evenly distribute professors who have already adopted the practice of giving individualized feedback. For other law schools, some instructors would indeed have to take on a heavier teaching burden. But even in a law school class of eighty students, it would probably take an instructor about forty additional hours to provide individualized feedback to students on their written work product. Moreover, professors who do not have even this much time to carve out of their schedules can provide individualized feedback through alternative methods, such as multiple-choice exams or teaching assistants.⁸¹ Given the large cost of law school tuition and

80. Emily Zimmerman, *What Do Law Students Want?: The Missing Piece of the Assessment Puzzle*, 42 RUTGERS L. J. 1, 5 (2010).

81. As suggested in Part I.A, *supra*, the extant literature on legal education includes numerous suggestions for efficiently providing enhanced feedback to students.

the still-challenging job market facing many new law school graduates, law schools owe their students at least this much.⁸²

More broadly, this article suggests that law schools should more systematically investigate when and how improved individualized feedback can affect student performance. The relative dearth of evidence on effective pedagogical techniques in law stands in significant contrast to the growing body of research empirically examining how and when different pedagogical approaches can improve learning outcomes in a variety of specific educational contexts. Leaders in legal education have begun to think more systematically about this issue, as evidenced by the ABA's new requirements for schools to develop learning outcomes and integrate formative feedback into their curriculum. But many questions remain unanswered, including the relative impact of different types of feedback, the marginal benefits of feedback in more than one class, and the potential for individualized feedback to reduce the broad tendency of minority students to underperform in law school even after controlling for their LSAT scores and undergraduate GPA. We hope this article helps spur more systematic empirical research into these questions and the many other unanswered issues surrounding the optimal approach to training future lawyers.

Conclusion

Exploiting a natural experiment at the University of Minnesota Law School from 2011 to 2015, this article empirically demonstrates that students who receive individualized feedback in a single first-year law school class outperform students who do not in class that they take jointly. This result rigorously confirms what much of the extant literature suggests—that providing students with individualized feedback designed to help them learn does indeed promote learning in law school. But it also does much more than that. In particular, it shows that the positive impacts of individualized, formative feedback extend well beyond the classroom in which that feedback is given, helping students compete in all their other law school classes. For those who have long preached that the core mission of law school classes is to teach students to “think like lawyers,” this result should not be shocking. To the extent that students in a single law school class are better taught how to think like a lawyer because they received individualized, formative feedback from their instructors, there is every reason to expect that this effect would extend to the remainder of law students' classes. The results also suggest that the benefits of individualized feedback are particularly acute for students at the University of Minnesota Law School who are in the bottom of their class or who arrive at law school with below-median LSAT scores.

82. Indeed, on the basis of this research, the University of Minnesota Law School has already revised its approach to constructing first-year sections and increased commitments among professors to provide individualized feedback. We hope that other law schools follow this trend.

We acknowledge, of course, that even the most carefully designed experimental studies cannot absolutely isolate cause and effect. In our study, for example, it is possible—although we think it unlikely—that some instructor-specific characteristic (other than clarity, which we tested) is responsible for driving the disparities that we found in students’ grades. Follow-up research in this area might shed more light on this possibility. Nonetheless, when considered in light of preexisting empirical research, our results provide convincing evidence of the value of incorporating feedback into the first-year curriculum. As such, we believe that law schools should systematically provide first-year law students with individualized feedback in at least one “core” doctrinal first-year class before final exams. As they do so, researchers and law schools should systematically investigate when and how improved individualized feedback can most powerfully affect student performance.