Does Practice Make Perfect?: An Empirical Examination of the Impact of Practice Essays on Essay Exam Performance

Andrea A. Curcio
0@0.com

Gregory Todd Jones
1@1.com

Tanya M. Washington
2@2.com

FLORIDA STATE UNIVERSITY LAW REVIEW

DOES PRACTICE MAKE PERFECT?
AN EMPIRICAL EXAMINATION OF THE IMPACT OF
PRACTICE ESSAYS ON ESSAY EXAM PERFORMANCE

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

ANDREA A. CURCIO, GREGORY TODD JONES &
TANYA M. WASHINGTON*

ABSTRACT

The recently published Best Practices and Carnegie Foundation studies challenge the legal academy to examine how we educate potential lawyers. This Article builds upon those studies by developing a cost-effective, duplicable empirical model that academics can use to measure the impact of various suggested teaching methodology improvements. Using that model, the Article discusses the authors' findings that, on average, first-year Civil Procedure law students who completed five practice essay writing exercises and received generalized postexercise feedback had higher raw scores on their Civil Procedure final essay exam questions than those who did not have the writing assignments. It also discusses the initially surprising findings that the most statistically significant benefit from the practice and feedback inured to students with above-the-median LSAT scores and above-the-median undergraduate GPAs and that the practice exercises had no impact upon students' grades in other courses. The Article discusses how these findings fit into the existing literature from other disciplines. It also explores how students' metacognitive skills may have influenced the study's outcome. The Article discusses how the incorporation of metacognitive skills training into the practice review sessions might result in a more across-the-board benefit from the exercises. It concludes with suggestions for future studies and for improving the original empirical model.

I.   INTRODUCTION

Two recent seminal studies challenge the legal academy to examine how we educate potential lawyers.[1] These studies lay the theoretical groundwork for thinking about the effectiveness of law school teaching methodologies. As these studies note, most law professors teach first-year courses via the case method[2] or the problem method[3]

---

1.   ROY STUCKEY ET AL., BEST PRACTICES FOR LEGAL EDUCATION (2007) (reporting the findings of a six-year project involving law faculty from throughout the country who sought to develop a blueprint for improving legal education); WILLIAM M. SULLIVAN ET AL., EDUCATING LAWYERS: PREPARATION FOR THE PROFESSION OF LAW (2007) (discussing the conclusions of Carnegie Foundation employees and grantees who spent years studying and analyzing the current education methods used by law schools and law professors across the country).

2.   As recognized by Professor Weaver, the term "case method" is commonly used to describe the process of having students read judicial opinions and then engaging them in a question-and-answer Socratic-style dialogue about those decisions and their implications. Russell L. Weaver, *Langdell's Legacy: Living with the Case Method*, 36 VILL. L. REV. 517, 518 (1991). For those who are unfamiliar with what is meant by the term "Socratic method" as it is used in context of law school classes, see STUCKEY ET AL., *supra* note 1, at 207-11. Most law faculty continue to use the case method to teach first-year law classes. *See* Steven I. Friedland, *How We Teach: A Survey of Teaching Techniques in American Law Schools*, 20 SEATTLE U. L. REV. 1, 28 (1996) (noting that out of 383 first-year professors who were surveyed, "370 or ninety-seven percent[ ] used the Socratic method at least some of the time in first year classes," with 71% of the respondents stating they used it most of the time or often); *see also* STUCKEY ET AL., *supra* note 1, at 133 (asserting that "[t]he main impediment to improving law school teaching is the enduring over reliance on the Socratic dialogue and case method."); SULLIVAN ET AL., *supra* note 1, at 47-78 (discussing the pros and cons of the traditional case-dialogue method).

3.   There has been a relatively recent move toward using the problem method or a combined case and problem method in first-year and other substantive courses. *See, e.g.*, Stephen J. Shapiro, *Teaching First-Year Civil Procedure and Other Introductory Courses by the Problem Method*, 34 CREIGHTON L. REV. 245, 246-54 (2000) (discussing why law professors should consider the problem method rather than the case method in first-year courses); *see also* Myron Moskovitz, *Beyond the Case Method: It's Time to Teach with Prob-*

without ever examining whether these methods adequately prepare students to practice law.[4] Using these largely empirically unexamined methodologies,[5] law professors endeavor to teach first-year students to identify issues, break legal rules into their component parts, understand the policy underlying legal rules, apply existing rules to new facts, synthesize a large number of rules, and extrapolate from existing rules and policies to create and justify new legal rules.[6] When they finally measure whether students have acquired the skills they sought to teach, they do not actually measure achievement of particular skill sets except in a comparative sense—they measure students against each other via a grading curve.[7] They do

---

lems, 42 J. LEGAL EDUC. 241 (1992) (arguing for use of the problem method in substantive law school courses).

4. *See generally* STUCKEY ET AL., *supra* note 1; SULLIVAN ET AL., *supra* note 1.

5. *See* James R. P. Ogloff et al., *More Than "Learning to Think Like a Lawyer:" The Empirical Research on Legal Education*, 34 CREIGHTON L. REV. 73, 181-83 (2000) (noting that there have been few attempts to empirically measure whether law school teaching methods are effective and most of the empirical measurements look at student satisfaction rather than learning outcome assessments). Our research indicates that in the last fifty years, there have been only a few empirical examinations of law school classroom teaching methodology. *See, e.g.*, Edward L. Kimball & Larry C. Farmer, *Comparative Results of Teaching Evidence Three Ways*, 30 J. LEGAL. EDUC. 196, 199-202 (1979) (finding little difference in learning outcomes between the case method, problem method or self-instruction method); *see also* Paul F. Teich, *Research on American Law Teaching: Is There a Case Against the Case System?*, 36 J. LEGAL EDUC. 167, 174-79 (1986) (describing nine studies of law teaching methods, most of them with undergraduate law classes, virtually all of which concluded that, in terms of various teaching methodologies, the only methodologies that had any impact on learning outcomes were computer-assisted methodologies or those geared toward individualized instruction). There have been a few other studies that were not scientifically valid, but nonetheless showed interesting results. *See, e.g.*, John M. Burman, *Out-Of-Class Assignments as a Method of Teaching and Evaluating Law Students,* 42 J. LEGAL EDUC. 447, 456-57 (1992) (finding that out-of-class writing assignments resulted in the author's subjective perception that students performed better on the final exam after doing the assignments); Stephen J. Shapiro, *The Use and Effectiveness of Various Learning Materials in an Evidence Class,* 46 J. LEGAL EDUC. 101, 108 (1996) (finding that between casebook reading, hornbook reading, and CALI exercises, only the CALI exercises seemed to have an effect on student exam performance). There also has been a recent empirical study that sought to examine the development of law students' legal reasoning skills as they progress through law school but did not examine how specific teaching methods effected this skill development. *See* Stefan H. Krieger, *The Development of Legal Reasoning Skills in Law Students: An Empirical Study*, 56 J. LEGAL EDUC. 332 (2006).

6. *See* STUCKEY ET AL., *supra* note 1, at 21-22 (citing to Judith Wegner's description of what students learn in the first year: thinking like a lawyer, legal literacy, legal analysis, application, and to some extent synthesis and evaluation); Steven Friedland, *A Critical Inquiry into the Traditional Uses of Law School Evaluation,* 23 PACE L. REV. 147, 201-02 (2002) (describing the competencies legal educators seek to teach law students); Weaver, *supra* note 2, at 548-49 (discussing how the case method, along with Socratic dialogue, teaches students to analyze and dissect a case and examine legal rules and their implications).

7. For a discussion of the normalization of grades in law school courses, see STUCKEY ET AL., *supra* note 1, at 243-45; SULLIVAN ET AL., *supra* note 1, at 168-70; Barbara Glesner Fines, *Competition and the Curve*, 65 UMKC L. REV. 879 (1997); Douglas A. Henderson, *Uncivil Procedure: Ranking Law Students Among Their Peers*, 27 U. MICH. J.L. REFORM

not generally attempt to identify what skills and knowledge individual students have or have not acquired.[8] Nor do most law professors devise methods of empirically assessing their effectiveness in helping students acquire particular skill sets. To the extent law professors have conducted empirical research, the research generally has not attempted to isolate the impact of a methodology on a particular skill set, but rather has looked at whether a particular teaching methodology (e.g., case method versus problem method) is better at teaching multiple course objectives.[9]

Law professors are not alone in their failure to apply the same rigorous examination to their teaching methodology as they do to their other scholarly pursuits. Beginning in 1990, recognizing the lack of scholarly examination of the effectiveness of teaching methodologies, Ernest Boyer proposed a new paradigm for pedagogical scholarship that encompassed scholarly assessment of the effectiveness of teaching methodologies.[10] The idea of the "scholarship of teaching and learning" began to be explored at the undergraduate level in the 1990s, but exactly what that scholarship should encompass and how it should be evaluated was an open question.[11] In 1999, Pat Hutchings and Lee S. Shulman proposed that the developing area of teaching scholarship should entail question-asking, inquiry, and investigation of student learning[12] that lays out the "vision, design, enactment outcomes and analysis—in a manner susceptible to critical review by the teacher's professional peers and amenable to productive employment in future work by members of that same community."[13]

The scholarship of teaching and learning has begun to take hold in the undergraduate arena where some scholars have worked to measure things such as the impact of practice tests on student learn-

---

399 (1994); Deborah Waire Post, *Power and the Morality of Grading—A Case Study and a Few Critical Thoughts on Grade Normalization*, 65 UMKC L. REV. 777 (1997).

    8. S*ee, e.g.,* STUCKEY ET AL., *supra* note 1, at 241-45 (arguing that law schools should move from norm-referenced to criteria-referenced assessments because criteria-referenced assessments are more reliable, valid, and useful to student learning); SULLIVAN ET AL., *supra* note 1, at 168-70.

    9. Teich, *supra* note 5, at 174; *see also* sources cited *supra* note 5 (describing the few studies that look at the effectiveness of teaching methodology in law school classrooms).

    10. *See generally* ERNEST L. BOYER, SCHOLARSHIP RECONSIDERED: PRIORITIES OF THE PROFESSORIATE (1990) (urging universities to conceptualize research and scholarship more broadly to encompass scholarship of teaching and learning).

    11. *See, e.g.*, CHARLES E. GLASSICK ET AL., SCHOLARSHIP ASSESSED: EVALUATION OF THE PROFESSORIATE (1997) (proposing guidelines for those seeking to engage in the scholarship of teaching and learning and for those seeking to assess the quality of scholarly work examining teaching and learning).

    12. Pat Hutchings & Lee S. Shulman, *The Scholarship of Teaching: New Elaborations, New Developments, in* LEARNING FROM CHANGE 47, 48 (Deborah DeZure ed., 2000).

    13. *Id.* (quoting Lee S. Shulman, *Course Anatomy: The Dissection and Analysis of Knowledge Through Teaching, in* THE COURSE PORTFOLIO 6 (Pat Hutchings ed., 1998)).

ing[14] and the effect of metacognition on student comprehension and performance.[15] However, even in the undergraduate context, the study of the effectiveness of various teaching methodologies on the acquisition of a particular skill set has been somewhat limited. For example, we have found no work attempting to measure empirically the impact of practice writing exercises, combined with feedback, on student essay exam performance. To begin to fill this void, we designed a cost-effective study that could be peer-reviewed and duplicated.[16] In the spring of 2006, we undertook a relatively simple empirical study which examined whether required practice writing assignments, combined with self-assessment, peer-assessment, and some faculty input had any discernible impact on first-year law students' ability to break a legal rule into its component parts and analyze and apply facts to each of the rule's elements in the context of an essay examination. We chose to study the impact of practice writing exercises on these particular skills for a number of reasons. First, others in the academy had suggested that practice writing exercises would improve student performance,[17] but none had empirically examined this hypothesis. Second, we believed that breaking a rule into its component parts and learning how to analyze facts in light of a rule's elements would lead to clearer thinking, organization, and writing and thus had potential implications for teaching in other disciplines. Third, we felt that the skills involved in breaking a rule into its component parts and applying facts to each of the rule's elements could be taught explicitly. Finally, in our experience, we have found that complex factual analysis is one of the most elusive skills for many first-year law students.

This Article expands upon an earlier article in which we briefly described the impact of five required writing exercises accompanied

---

14. *See infra* text accompanying notes 33-40 (discussing studies involving the impact of practice tests on student exam performance).

15. *See* Michael Hunter Schwartz, *Teaching Law Students to Be Self-Regulated Learners,* 2003 MICH. ST. DCL L. REV. 447, 474-84 (discussing various studies outside of legal education that compare the performance of expert and novice self-regulators). In the realm of U.S. legal education, only a few empirical studies have attempted to measure the impact of metacognition on students' performance, and those studies focused on students' ability to read and analyze cases and law review articles. *Id.* at 473-74. For a more detailed description of some of the studies involving metacognition, see *infra* text accompanying notes 42-66.

16. This comports with the recommendation of Hutchings and Shulman, *supra* text accompanying notes 12-13, who suggest that the scholarship of teaching and learning should entail investigation and analysis that is both subjected to critical peer review and capable of duplication by others. For a description of the methodology used in this study, see *infra* Part IV.

17. *See, e.g.*, Mary Beth Beazley, *Better Writing, Better Thinking: Using Legal Writing Pedagogy in the "Casebook" Classroom (Without Grading Papers)*, 10 J. LEGAL WRITING INST. 23, 72-74 (2004); Burman, *supra* note 5, at 452-53, 455, 457 (discussing author's subjective impression that practice writing exercises improved his students' essay exam performance).

by generalized feedback on first-year law students' Civil Procedure essay exam performance.[18] This Article explores further our findings that: 1) although students receiving the writing interventions had higher raw essay exam scores on average, the most statistically significant benefit from the interventions seemed to inure to those students with above-the-median LSAT scores and above-the-median undergraduate grade point averages (UGPA); and 2) that the writing interventions did not seem to have a measurable spillover effect on grades in other first-year courses. It discusses how these empirical findings fit into the existing literature and how metacognition and speededness[19] may help explain the study's results. This Article also details the study's design so that those wishing to duplicate or improve upon the study may do so. Finally, it explores student reaction to the writing interventions.

Part II of this Article briefly reviews the existing literature on law school teaching and testing, the value of practice tests, the role metacognition plays in reading comprehension and student performance, and provides the background information necessary to understand some potential explanations for our findings. Part III describes how we integrated the writing interventions into a substantive first-year course and the method we used to test whether these writing interventions had any discernible impact on student learning outcomes. Part IV describes the empirical findings that although, on average, students in the intervention class had higher raw essay exam scores, the most statistically significant difference in scores was found when comparing intervention and nonintervention students with above-the-median LSAT scores and above-the-median UGPAs. It also discusses the findings that the interventions did not have a measurable spillover effect in other courses. Using the existing literature as background, Part V suggests some possible reasons for the study's findings. Part V also summarizes students' reaction to the writing interventions and feedback. Part VI discusses ways the study can be used, and improved upon, by others. It also suggests further areas ripe for potential inquiry. Finally, the Article concludes by encouraging others to use this model as a building block for further empirical study of various teaching methodologies.

---

18. *See* Andrea A. Curcio et al., *Developing an Empirical Model to Test Whether Required Writing Exercises or Other Changes in Large-Section Law Class Teaching Methodologies Result in Improved Exam Performance*, 57 J. LEGAL EDUC. 195 (2007).

19. "Speededness" and "test-taking speed" are related variables. For a detailed discussion of the relation between these variables, 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 (2004). In this Article, we use both terms to describe the effect of time limits on exam performance.

II. Background Literature

In order to understand this study, its results, and the significance of those results, it is important to review the existing literature on empirical assessments of law school teaching methodologies, the impact of practice tests on student performance, the role of metacognition in student performance, and basic information about what the LSAT purports to measure. This Part provides that review.

A. *Assessments of Law School Teaching Methodologies*

For over one hundred years, the predominant method of law school teaching has been the Langdellian casebook method. Students read and analyze cases and, in class, students respond to questions about the cases and questions that force students to look at the implications of the cases in different factual settings and in the context of other legal rules and social policy issues.[20]

Many suggestions have been made to improve our teaching including: using multiple assessments;[21] giving meaningful opportunities to practice skills we want students to learn;[22] being more explicit about the skills we want students to learn;[23] allowing students an opportunity for self-assessment and peer-assessment;[24] and providing guidelines to help students self-assess.[25]

However, scholars have done little empirical research to examine whether these suggested changes result in improved student learning outcomes.[26] The empirical work on law school teaching methodology has thus far been mainly limited to studying whether the case method or problem method result in better overall learning outcomes,[27] how better to teach students to read cases,[28] and how law

---

20. *See* sources cited *supra* note 2 and accompanying text.

21. Gregory S. Munro, Outcomes Assessment for Law Schools 74-75 (2000); Stuckey et al., *supra* note 1, at 255-60; Friedland, *supra* note 6, at 188.

22. *See* Stuckey et al., *supra* note 1, at 255-60 (suggesting professors use multiple formative and summative assessments throughout the semester); Friedland, *supra* note 6, at 208; Phillip C. Kissam, *Law School Examination*s, 42 Vand. L. Rev. 433, 473 (1989).

23. Beazley, *supra* note 17, at 72-74; Sophie M. Sparrow, *Describing the Ball: Improve Teaching By Using Rubrics—Explicit Grading Criteria*, 2004 Mich. St. L. Rev 1 *passim*.

24. Stuckey et al., *supra* note 1, at 254; Friedland, *supra* note 6, at 209; Greg Sergienko, *New Modes of Assessment*, 38 San Diego L. Rev. 463, 482 (2001).

25. Beazley, *supra* note 17, at 72-74; *see also* Stuckey et al., *supra* note 1, at 254; Mary A. Lundeberg, *Metacognitive Aspects of Reading Comprehension: Studying Understanding in Legal Case Analysis*, 22 Reading Res. Q. 407, 428 (1987) (finding that explicitly teaching strategies for reading and understanding legal opinions significantly improved comprehension of novice law students).

26. At least one professor has noted that from a subjective standpoint, out-of-class writing assignments seem to improve final exam performance. *See* Burman, *supra* note 5, at 452-53, 455, 457.

27. *See supra* note 5 (discussing studies of teaching methodologies in law classes).

28. *See* Lundeberg, *supra* note 25; *see also* Robin A. Boyle, *Employing Active-Learning Techniques and Metacognition in Law School: Shifting Energy from Professor to Student*,

students' legal reasoning skills develop during their three years in law school.[29] In this study, we begin the development of an empirical model to examine teaching methodology as it relates to the acquisition of a specific skill set. Specifically, we looked at students' ability to break a legal rule into its component parts and then analyze and apply complex facts to each of the rule's elements, incorporating many of the empirically unexamined suggestions for improving student learning.[30] We then attempted to measure whether, in fact, these interventions had any impact on student learning outcomes and whether our findings were consistent with existing studies.

B. *Assessing the Impact of Practice Tests*[31]

One suggestion for changing how we teach in order to improve student learning outcomes is to give students more opportunities to practice the skills we want them to learn.[32] One way to do this is via practice tests. In other disciplines, empirical studies have found that practice tests may contribute to improved performance on actual tests.[33] Practice tests were most effective in improving performance on regular exams if the practice exams closely matched the format and difficulty of the actual exams.[34] Reasons postulated to explain

---

81 U. DET. MERCY L. REV. 1, 15-17 (2003) (discussing other studies examining reading comprehension of legal opinions and law review articles); Schwartz, *supra* note 15, at 473-74 (discussing empirical research on the effect metacognition plays in legal reading).

   29. *See* Krieger, *supra* note 5.

   30. We gave students in the writing intervention cohort multiple meaningful opportunities to practice skills we wanted them to learn. We were explicit about the skills we wanted them to learn. We allowed them an opportunity for self-assessment and peer-assessment and we provided them with guidelines to help them self-assess. We also provided students with some individualized feedback. These were all suggestions other scholars had for improving student learning. *See supra* text accompanying notes 21-25.

   31. Although the writing interventions were not practice tests per se, they were similar to the kind of single-issue essay questions students might see on a final exam. Thus, the literature on the value of practice tests is instructive for this writing intervention study.

   32. *See* STUCKEY ET AL., *supra* note 1, at 259-61; Friedland, *supra* note 6, at 208; Kissam, *supra* note 22, at 473.

   33. *See* William R. Balch, *Practice Versus Review Exams and Final Exam Performance*, 25 TEACHING PSYCHOL. 181, 182-83 (1998) (practice tests improved final exam performance for students at all levels of class standing); John A. Gretes & Michael Green, *Improving Undergraduate Learning with Computer-Assisted Assessment*, 33 J. RES. ON COMPUTING EDUC. 46, 48 (2000) (finding computerized practice tests improved final test performance); Margaret K. Snooks, *Using Practice Tests on a Regular Basis to Improve Student Learning*, 100 NEW DIRECTIONS FOR TEACHING & LEARNING 109, 110 (2004) (finding students got better grades when given practice tests and that students found the practice tests helpful to their review of the material). *But see* Linda Bol & Douglas J. Hacker, *A Comparison of the Effects of Practice Tests and Traditional Review on Performance and Calibration*, 69 J. EXPERIMENTAL EDUC. 133, 134 (2001) (noting that some studies have found that the gains from practice tests may not be appreciatively greater than the gains that could be made through regular instruction).

   34. Renee Oliver & Robert L. Williams, *Direct and Indirect Effects of Completion Versus Accuracy Contingencies on Practice-Exam and Actual-Exam Performance*, 14 J. BEHAV.

why practice tests improve test performance include "greater famili-arity with the kinds of test items, improved test-taking skills, greater confidence in testing, and higher levels of testing sophistication."[35] However, in some cases, practice tests may not make a difference in student exam performance. For example, one study found that be-tween students given a practice test with a review session going over practice test questions, and students given a traditional review ses-sion, the practice test cohort did not perform better on exams;[36] in fact, students taking the practice tests actually did worse on the mid-term multiple choice exam[37] and the practice had no impact on stu-dent performance on the short-answer essay questions.[38]

Because we used the LSAT as one of the variables to measure whether student exam performance was related to the interventions or some other cause, we searched for literature which used the SAT as a variable in looking at the impact of practice tests on under-graduate performance. We only found one study that looked at the correlation between practice tests and SAT scores.[39] That study found that for multiple-choice exams, computerized practice tests improved grades on average for students who took the practice tests regardless of SAT scores, although more students with high SAT scores took ad-vantage of the practice test opportunities.[40]

Thus, the literature empirically measuring the impact of prac-tice tests and exercises on exam performance generally supports the proposition that if the practice tests are similar to the actual test, practice improves performance of all students, regardless of SAT scores. However, these studies focused on multiple-choice and short

EDUC. 141, 142 (2005) (citing to studies which indicate that practice exams which are simi-lar to regular exams may improve students' exam performance).

35. Bol & Hacker, *supra* note 33, at 134 (citing ANNE ANASTASI, PSYCHOLOGICAL TESTING 23-47 (6th ed. 1988)).

36. *Id.* at 140.

37. *Id.* Professors Bol and Hacker postulate that this result may be due to how stu-dents used the practice tests as a review tool. "If the students expected identical or nearly identical items on the exam, they may have focused their study efforts too narrowly on the same content that appeared on the practice items." *Id.* at 147. Additionally, they suggest that the results may be explained because the review session for the group without the practice exam may have been a more comprehensive overview of the material rather than the more limited focus involved in reviewing answers to the practice test questions. *Id.*

38. *Id.* at 140. We searched for other studies involving an empirical examination of the impact of practice essay assignments on exam performance and did not find any such studies. This may be because most of the work in this area is in undergraduate classes, many of which test via multiple choice exams. However, we realize that the failure to find studies dealing with empirical assessments of essay practice tests or assignments on exam performance may also be due to our limited familiarity with how to search the social sci-ence literature. We did find one article in which a legal educator discussed his subjective perception that out-of-class writing assignments led to better performance on final exam essay questions. *See* Burman, *supra* note 5, at 452-53, 455, 457.

39. Gretes & Green, *supra* note 33.

40. *Id.* at 48, 50-51.

answer question tests. We found no empirical studies that looked at the impact of practice writing exercises on essay exam performance.[41] Part V discusses how our results correspond to, differ from, and expand upon these studies' findings.

C. *Assessing the Relationship Between Metacognition and Student Comprehension and Performance*

Metacognition is the term used to describe understanding what one knows and how to learn best what one does not know.[42] It involves both knowledge of cognition and knowledge of how to self-regulate one's own learning.[43] Knowledge of cognition includes "knowledge about ourselves as learners and what factors influence our performance."[44] It also includes knowledge about learning strategies such as "taking notes, slowing down for important information, skimming unimportant information, using mnemonics, summarizing main ideas, and periodic self-testing."[45] Metacognition also encompasses the regulation of cognition: choosing appropriate learning strategies, allocation of resources and time, self-testing, self-monitoring and knowing when and how to adjust learning strategies when necessary.[46] Metacognition "help[s] learners use their attentional resources more efficiently, process information at a deeper level and monitor their performance more accurately."[47]

---

41. One law professor has written about his subjective observations that practice writing exercises improve exam performance. *See* Burman, *supra* note 5, 452-53, 455, 457.

42. *See* Anthony S. Niedwiecki, *Lawyers and Learning: A Metacognitive Approach to Legal Education*, 13 WIDENER L. REV. 33, 35 (2006) (noting that, "[g]enerally, metacognition refers to having both awareness and control over one's learning and thinking"); Christine M. Venter, *Analyze This: Using Taxonomies to "Scaffold" Students' Legal Thinking and Writing Skills*, 57 MERCER L. REV. 621, 636 (2006) (metacognition is consciously "thinking about one's own thinking strategies").

43. Niedwiecki, *supra* note 42, at 43-44.

44. Gregory Schraw & David Brooks, Improving College Teaching Using an Interactive, Compensatory Model of Learning 6 (Sept. 5, 2006) (on file with author) (summarizing findings of Ann Brown, *Metacognition, Executive Control, Self-Regulation, and Other More Mysterious Mechanisms*, *in* METACOGNTION, MOTIVATION AND UNDERSTANDING 65 (Franz E. Weinert & Rainer H. Kluwe eds., 1987) and Janis E. Jacobs & Scott G. Paris, *Children's Metacognition About Reading: Issues in Definition, Measurement and Instruction*, 22 EDUC. PSYCHOLOGIST 255 (1987)).

45. *Id.*; *see also* Linda Baker, *Metacognition, Comprehension Monitoring, and the Adult Reader*, 1 EDUC. PSYCHOL. REV. 3, 6-7 (1989) (describing metacognitive strategies used by expert readers).

46. Schraw & Brooks, *supra* note 44, at 6.

47. John L. Nietfeld et al., *Metacognitive Monitoring Accuracy and Student Performance in the Postsecondary Classroom*, 74 J. EXPERIMENTAL EDUC. 7, 9 (2005); *see also* Schwartz, *supra* note 15, at 474-77 (discussing studies which indicate that self-regulation impacts students' performance).

Metacognition also plays a significant role in reading comprehension.[48] Strong readers employ a variety of metacognitive strategies to increase and regulate comprehension.[49] For example, good readers repair comprehension problems by employing strategies such as "rereading the text, summarizing, making inferences or consulting outside help."[50] Thus, metacognitive skills are employed both in monitoring and in repairing reading comprehension. Better students report using more and different kinds of criteria for evaluating their understanding of what they have read.[51] Undergraduate students who reported monitoring their comprehension and dealing with comprehension failures earned higher GPAs than students who did not.[52] Likewise, undergraduate students who used a higher number of different criteria for evaluating their comprehension earned higher course grades.[53]

Students with highly developed metacognitive skills are able to use various strategies to increase comprehension. They also are able to use these skills to help them maximize their exam performance. One study postulated that these skills have a "substantial additional value on top of intelligence in explaining academic performance."[54] The relationship between metacognitive skills and academic performance likely is related both to how metacognitive skills improve comprehension and how they impact students' study strategies.[55] For example, metacognitive skills come into play when students are asked to predict how they will perform on a test. Studies link the ability to predict exam performance to academic performance and show that those with strong predictive ability have, on average, higher UGPAs.[56] One reason for the correlation between predictive ability and academic performance may be that students who are bet-

---

48. *See generally* Baker, *supra* note 45 (providing an overview and synthesis of the literature on the role of metacognition and adult readers' comprehension); Lundeberg, *supra* note 25 (discussing the role of metacognition in reading legal opinions).

49. Baker, *supra* note 45, at 6-9.

50. Peter Dewitz, *Legal Education: A Problem of Learning from Text*, 23 N.Y.U. REV. L. & SOC. CHANGE 225, 229 (1997).

51. Baker, *supra* note 45, at 8-9.

52. *Id.* at 9; *see also* Schwartz, *supra* note 15, at 476 (discussing studies in which students with strong self-regulation skills earned higher grades than those with lower self-regulation skills).

53. Baker, *supra* note 45, at 10.

54. Alexander Minnaert, *Can Metacognition Compensate for Intelligence in the First Year of Belgian Higher Education?*, 36 PSYCHOLOGICA BELGICA 227, 239-40 (1996). Interestingly, Professor Minnaert notes that there is a low correlation between intelligence as measured by an intelligence test and metacognitive knowledge. *Id.* at 236.

55. Niedwiecki, *supra* note 42, at 45 (noting "[s]tudents possessing the ability to accurately distinguish between what has already been learned and mastered from what is yet to be learned have a far greater advantage, as they can be more strategic and effective learners," especially in law school where students have to absorb "a great deal of information in a limited amount of time").

56. Balch, *supra* note 33, at 181; Nietfeld et al., *supra* note 47, at 19.

ter able to predict how they will perform are better able to target their efforts when it comes to studying.[57] In fact, one metacognitive strategy that has been isolated and linked to superior test performance is successfully verbalizing what the student expects the test questions to be.[58]

Studies have shown that metacognitive skills can be taught,[59] although newly acquired strategies do not readily transfer to new tasks or unfamiliar domains.[60] One study indicates that students can improve, in a modest way, the regulation of their behavior such that the reflective skills taught can impact performance in courses other than the ones in which the skills were explicitly being used.[61] However, this must be done through specific strategies directed at improving students' metacognitive abilities. "Merely prompting students to think about their performance [on practice tests] is likely to be too passive of an attempt to alter monitoring accuracy" and improve metacognitive skills.[62] Improving metacognitive skills, and thus improving academic performance, requires practice, feedback and employing strategies on a consistent, intensive, and explicit basis.[63]

Almost twenty years ago, Professor Wangerin suggested that law professors teach students metacognitive strategies[64] and a few professors have built upon his work in designing classes or suggesting different teaching strategies.[65] There have been only a few studies involving metacognition and legal learning and those have been

---

57. As Professor Nietfeld noted, "poor students understand they are poor students but may not know where to target their efforts to improve, whereas better students may tend to be more strategic and aware of where they need to expend their efforts toward improvement." Nietfeld et al., *supra* note 47, at 24; *see also* Niedwiecki, *supra* note 42, at 45 (noting that strong metacognitive skills help students better focus their study time and energy).

58. Baker, *supra* note 45, at 9-10.

59. *See* Lundeberg, *supra* note 25, at 428-29; Chris Masui & Erik De Corte, *Learning to Reflect and to Attribute Constructively as Basic Components of Self-Regulated Learning*, 75 BRIT. J. EDUC. PSYCHOL. 351, 364-66 (2005); Barry J. Zimmerman & Andrew S. Paulsen, *Self-Monitoring During Collegiate Studying: An Invaluable Tool for Academic Self-Regulation*, 63 NEW DIRECTIONS FOR TEACHING AND LEARNING 13, 14, 22-23 (Paul R. Pintrich ed., 1995), (concluding that self-regulatory skills are "highly predictive of students' academic success[] and that these skills can be taught").

60. Schraw & Brooks, *supra* note 44, at 5.

61. *See generally* Masui & De Corte, *supra* note 59 (discussing a study in which students were trained in the metacognitive skills of reflection and attribution and the impact of this training on students' academic achievement).

62. Nietfeld et al., *supra* note 47, at 22.

63. *Id.*

64. Paul T. Wangerin, *Learning Strategies for Law Students*, 52 ALB. L. REV. 471, 478-79 (1988) ("[E]ducators should provide students with the tools to understand, monitor, and adapt their study activities to accomplish particular academic goals.").

65. *See, e.g.*, Boyle, *supra* note 28, at 18-28; Paula Lustbader, *Construction Sites, Building Types, and Bridging Gaps: A Cognitive Theory of the Learning Progression of Law Students*, 33 WILLAMETTE L. REV. 315, 339, 344-46, 349-50, 352 (1997); Niedwiecki, *supra* note 42, at 62-68; Schwartz, *supra* note 15, at 484-505; Venter, *supra* note 42, at 638-42.

confined to the use of metacognitive strategies to increase case law reading comprehension.[66] In all candor, we were unfamiliar with how metacognition may affect learning when we designed and implemented our study. Only as we sought an explanation for our results did we begin to understand the role metacognition may have played in how students were able to use the information they acquired from the writing interventions. In Part V, we discuss our hypothesis that students' metacognitive abilities may have affected the study's outcome, and in Part VI, we discuss ways future studies could test this hypothesis.

D. Overview of the Law School Admission Test

Our results, as set out in Part IV, found a correlation between LSAT scores and the impact of the writing interventions on raw score essay exam performance. To contextualize these findings it is important to consider both what the LSAT measures and what it does not measure. The LSAT is a half-day, standardized, wholly multiple-choice exam required for admission to most U.S. and Canadian law schools.[67] The Law School Admission Council (LSAC), author of the LSAT, describes the exam as measuring "the reading and comprehension of complex texts with accuracy and insight; the organization and management of information and the ability to draw reasonable inferences from it; the ability to think critically; and the analysis and evaluation of the reasoning and arguments of others."[68]

The LSAT score is one of the primary law school admission criteria because it is thought to have predictive value in terms of poten-

---

66. *See* Lundeberg, *supra* note 25; *see also* Boyle, *supra* note 28, at 15-17 (discussing other studies examining reading comprehension of legal opinions and law review articles); Leah M. Christensen, *The Psychology Behind Case Briefing: A Powerful Cognitive Schema*, 29 CAMPBELL L. REV. 5, 6 n.6 (2006) (noting that the author is currently involved in an empirical study of how law students and lawyers read cases).

67. There is a writing sample component of the LSAT but the writing sample is not considered in the numerical LSAT score.

68. LSAC/LSDAS REGISTRATION AND INFORMATION BOOK 1 (2007). An LSAC study claims the LSAT measures reading comprehension by requiring test takers to "read carefully and accurately, to determine the relationships among the various parts of the passage, and to draw reasonable inferences from the material in the passage." Kenneth M. Wilson & Donald E. Powers, *Factors in Performance on the Law School Admission Test* 1 (LSAC, Statistical Rep. No. 93-04, 1994). The same study suggests the test measures logical reasoning via questions that require "the examinee to read and comprehend the argument or the reasoning contained in a short passage, and then answer one or two questions about it. The questions test a variety of abilities involved in reasoning logically and critically []including, for example,[] drawing reasonable conclusions from given evidence or premises." *Id.* Finally, the Wilson and Powers study suggests the LSAT measures analytical reasoning ability through questions "designed to measure the ability to understand a structure of relationships and to draw conclusions about the structure. The examinee is asked to make deductions from a set of statements, rules, or conditions that describe relationships among entities such as persons, places, things, or events." *Id.*

tial law students' ability to succeed in law school.[69] However, the LSAT's predictive value has been questioned both because it accurately predicts grades for only a relatively small percentage of students[70] and because, given the limited skills it measures, it does not predict which students will become good lawyers.[71]

Even the LSAC recognizes the limits inherent in any predictive value the test may have. "The LSAT is designed to measure some, but certainly not all, of the mental and academic skills that are needed for successful law study. Within limits, it provides a reasonable assessment of these factors."[72] Some suggest that the LSAT's

---

69. *See* Henderson, *supra* note 19, at 986 n.51 (2004) (setting forth numerous studies that claim the LSAT is a better predictor than UGPA alone in determining law school success). *But see* Jeffrey S. Kinsler, *The LSAT Myth*, 20 ST. LOUIS U. PUB. L. REV. 393, 416 (2001) (finding that although the LSAT is claimed to be a better predictor of first-year law school grades than UGPA, at Marquette the opposite was true—UGPA was a better predictor of law school performance than LSAT score).

70. "The district court in *Grutter v. Bollinger* acknowledged that trial evidence indicated that 'the LSAT predicts law school grades rather poorly (with a correlation of only 10-20%) and that it does not predict success in the legal profession at all.' " Pamela Edwards, *The Shell Game: Who is Responsible for the Overuse of the LSAT in Law School Admissions?*, 80 ST. JOHN'S L. REV. 153, 161 (2006) (footnote omitted). The limited predictive value of the LSAT has been noted by others. *See, e.g.*, William C. Kidder, *The Rise of the Testocracy: An Essay on the LSAT, Conventional Wisdom, and the Dismantling of Diversity*, 9 TEX. J. WOMEN & L. 167, 187 (2000) (reporting that "the LSAT accounts for only 16% of the variation in first-year grades among students enrolled in ABA law schools"); Abiel Wong, Note, *"Boalt-ing" Opportunity?: Deconstructing Elite Norms in Law School Admissions*, 6 GEO. J. ON POVERTY L. & POL'Y 199, 227 (1999) ("The LSAT's correlation coefficient with first-year grades ranges from .01 to .62, depending on the law school, with a median correlation with .41. When the LSAT is used in conjunction with [U]GPA, predictive validity increases (ranging from .11 to .68), with a median correlation coefficient of .49.").

71. William C. Kidder, *Does the LSAT Mirror or Magnify Racial and Ethnic Differences in Educational Attainment?: A Study of Equally Achieving "Elite" College Students*, 89 CAL. L. REV. 1055, 1115-16 (2001) (arguing that despite many changes in format over the years, the LSAT's correlation to first-year averages has remained constant and that the consistent correlation level "casts doubt on the notion that the LSAT is somehow uniquely designed to capture the set of skills required for the study of law. Rather, the data suggest that any major norm-referenced 'aptitude' test could stand in for the LSAT and produce equivalent results."); *see also* Edwards, *supra* note 70, at 158 ("[T]he LSAT does not consider other attributes that a successful law student should have, 'such as motivation, perseverance, listening or speaking skills, or writing ability.' "); Phoebe A. Haddon & Deborah W. Post, *Misuse and Abuse of the LSAT: Making the Case for Alternative Evaluative Efforts and a Redefinition of Merit*, 80 ST. JOHN'S L. REV. 41, 53-54 (arguing that the LSAT and first-year exams do not purport to measure anything close to the wide range of skills necessary for minimally competent lawyering); Ian Weinstein, *Testing Multiple Intelligences: Comparing Evaluation by Simulation and Written Exam*, 8 CLINICAL L. REV. 247, 248-50 (2001) (noting that the LSAT has limited utility in law school grade prediction and that law school grades do not correlate to or predict success in the practice of law). Given the limited utility of the LSAT in predicting who will be a good lawyer, many have argued for minimizing the use of the LSAT as an admissions criteria because of its potential discriminatory impact. *See, e.g.,* Kidder, *supra*, at 1119-24; Vernellia R. Randall, *The Misuse of the LSAT: Discrimination Against Blacks and Other Minorities in Law School Admissions*, 80 ST. JOHN'S L. REV. 107, 139-43 (2006).

72. LAW SCH. ADMISSION COUNCIL, LSAC STATEMENT OF GOOD ADMISSION AND FINANCIAL AID PRACTICES 2 (2007), http://www.lsac.org/pdfs/2007-2008/Statementof

predictive ability actually stems, in part, from the fact that the LSAT tests a psychometrically distinguishable ability, test-taking speed.[73] Speededness is also a significant component of many first-year law school exams.[74]

Finally, LSAT scores can be improved if LSAT applicants take an LSAT preparation course.[75] Practice LSAT courses such as The Princeton Review aim to improve students' LSAT scores by helping them "master LSAT content, build skills with practice tests and learn proven test-taking strategies."[76] In fact, The Princeton Review company is so confident that its review courses will improve test-takers' LSAT scores that it offers a qualified money-back guarantee if review course takers either do not improve their LSAT score, or are unsatisfied with their score.[77]

In sum, to the extent the LSAT has predictive value for first-year grades, this value likely results from the fact that the LSAT attempts to measure the same narrow subset of skills considered to be the focus of most first-year law school exams,[78] because speededness is a significant variable that informs performance in both examination contexts,[79] and also because successful test-taking for first-year

---

GoodAdm2007.pdf; *see also* Jerry R. Parkinson, *Admissions After* Grutter, 35 U. Tol. L. Rev. 159, 163 (2003) (noting that the LSAC does not dispute that the LSAT actually only accurately predicts first-year performance for a relatively small percentage of first-year students, but instead the LSAC "simply—and correctly—points out that there is no other measurement that 'comes close to matching the predictive qualities of the LSAT.' ").

73. Henderson, *supra* note 19, at 979 (2004) (presenting strong empirical evidence that speededness informs student performance on both the LSAT and law school exams). Henderson notes, "[w]ithin the field of psychometrics, test-taking speed and reasoning ability are viewed as distinct, separate abilities with little or no correlation." *Id.*

74. "[T]he LSAT is a good predictor of first-year law school grades because the test's heavy emphasis on time constraints is indicative of the nature of first-year in-class exams. . . . Thus, a higher score on such exams is not an indication of superior knowledge or better preparation. Given this outcome, the academy may wish to consider whether rewarding 'speediness' on law school exams is desirable for training students to practice law."

Edwards, *supra* note 70, at 163-64.

75. Jay Rosner, an expert witness in *Grutter v. Bollinger*, testified that LSAT prep courses like those offered by The Princeton Review or Kaplan "generally improve one's LSAT score by approximately seven points." Grutter v. Bollinger, 137 F. Supp. 2d 821, 860 (E.D. Mich. 2001), *rev'd on other grounds*, 539 U.S. 306 (2003).

76. The Princeton Review, Law School and the LSAT, www.princetonreview.com/law/testprep (last visited Feb. 15, 2008).

77. The Princeton Review, Our Guarantee To You, http://www.princetonreview.com/law/testprep/testprep.asp?TPRPAGE=508&TYPE=LSAT-PREPARE (last visited Feb. 15, 2008).

78. *See supra* note 68 and accompanying text (discussing what the LSAT tests). "A traditional essay exam calls for the application of recalled law to a factual situation. This tests the ability to read, to identify facts as triggering the application of legal rules, and to write analysis." Sergienko, *supra* note 24, at 469.

79. "If law school testing methods were unspeeded (e.g., take-home exams and papers), then, all other factors being equal . . . candidates [of simi-

law exams and the LSAT is a learned skill. Notably, looking at LSAT score as the only variable, students with high LSAT scores who did not receive the writing interventions were no more likely to get high raw essay question scores than students with low LSAT scores.[80] However, once we added the writing intervention variable, we found that the intervention provided a significant benefit to those with above-the-median LSAT scores[81] and no discernible benefit for students with below-the-median LSAT scores.[82] The extent to which students' metacognitive skills and the comparative similarity of the skill sets that may produce high LSAT scores and strong law school exam performance explains this result will be discussed in more detail in Part V below.

III. STUDY DESIGN AND IMPLEMENTATION

In the spring of 2006, one of the coauthors of this Article, a professor at a second-tier regional law school, sought to improve her eighty first-year Civil Procedure students' legal analysis and writing skills through a series of required writing exercises accompanied by various forms of feedback. To test whether the exercises and feedback had any discernible impact, the professor recruited the other two coauthors of this Article, a professor who taught a fifty-five student, first-year Civil Procedure class, and a faculty research fellow and adjunct law professor. The three collaborated to develop a way to measure the impact of the writing exercises. In designing the study, they designated the professor using the writing exercises as the "intervention professor" and the other as the "nonintervention professor."[83]

The nonintervention professor taught using a combination of the traditional case dialogue and the problem method.[84] Although she emphasized the need for students to apply the facts in any analysis,

---

lar reasoning ability] would tend to perform at the same level and the LSAT will have little or no predictive ability. However, if in-class exams with strict time limits are the dominant testing method, then the ordering of test-takers will tend to track the ordering of the LSAT, thus driving up the LSAT's predictive validity."

Henderson, *supra* note 19, at 1032-33 (footnote omitted).

    80. *See infra* Table 3.

    81. *See infra* Table 4.

    82. *See infra* Table 5.

    83. We used different textbooks but both supplemented our texts with materials in LEWIS A. GROSSMAN & ROBERT G. VAUGHN, A DOCUMENTARY COMPANION TO *A CIVIL ACTION* (rev. ed. 2002). This book contains pleadings, briefs, motions and other materials from a complex toxic tort case. The book's authors use the court documents as the basis for various questions about substantive legal procedure.

    84. *See supra* notes 2-3 (discussing case dialogue and problem method of teaching substantive law classes).

she did this through oral case analysis and hypothetical in-class problems.

The intervention professor also taught using the case dialogue and problem method. Additionally, she required her students to write five three-page, take-home papers. The papers were designed to help students learn how to break a legal rule into its component parts, analyze and apply facts to each of the rule's elements, and to make arguments on both sides. As discussed in more detail below, the intervention professor attempted to incorporate other professors' suggestions of: using multiple assessments, giving students meaningful opportunities to practice specific skills, telling the students explicitly the skills they should be learning and employing, allowing students an opportunity for self-assessment and peer assessment, and providing guidelines to help students self-assess.[85]

Two weeks into the spring semester, the intervention professor gave her students an initial single issue "practice" paper, which was worth one raw score point toward the final grade.[86] After the students turned the papers in, the intervention professor read approximately ten papers to get a sense of common errors and issues. Before assigning the next paper, she reviewed the IRAC formula (issue, rule, analysis, and conclusion)[87] with her class. She also gave students general feedback on common problems she saw in the papers she had read and discussed how to avoid those problems in future papers.

After that, the intervention professor assigned three additional three-page papers—one paper every two weeks. Students had a week to complete each paper. Each paper was a single-issue essay question involving a legal rule that had been discussed in class.[88] Students received four raw score points toward their final grade for each paper if

---

85. *See supra* notes 21-25 (discussing suggestions from various law faculty members regarding ways to improve law teaching).

86. This paper was a half-page factual scenario that required students to apply Federal Rule of Civil Procedure 20. Students needed to use the facts to discuss whether the parties' claims arose from "the same transaction, occurrence, or series of transactions or occurrences" and had a "question of law or fact common to all these persons." FED. R. CIV. P. 20(a).

87. "Professor Philip C. Kissam identified four intellectual functions tested, on the surface, in a blue book exam: (1) issue spotting; (2) identification of relevant legal authority; (3) application of legal authority to facts; and (4) organization of material." Adam G. Todd, *Exam Writing as Legal Writing: Teaching and Critiquing Law School Examination Discourse*, 76 TEMP. L. REV. 69, 72-73 (2003) (citing Kissam, *supra* note 22, at 466). This is commonly called the IRAC formula.

88. For all papers, students were given a one-page factual scenario. The first paper involved a Federal Rule of Civil Procedure 13 compulsory counter-claim issue. It required analysis of whether a claim arose out of the same transaction or occurrence or series of transactions or occurrences. The second paper required application of the component parts of Federal Rule of Civil Procedure 19(a) as students analyzed whether a party was a "necessary party." The third paper involved a work product issue.

they submitted it by the deadline. No credit was given for late papers. Students knew that one of the three papers would be graded with a minimum of four and a maximum of eight raw score points but they did not know which paper would be graded. The intervention professor hoped that this would motivate the students to put effort into every paper. Throughout the semester, the essay question assignments became progressively more difficult, either in terms of the legal rule or in terms of more complex and nuanced fact patterns.

After the students had turned in each paper, the students were given an annotated model answer designed to help them understand the writing and analysis process by noting things such as: "this paragraph illustrates how you use the facts to argue both sides" or "this is an example of breaking a rule into its component parts and analyzing each part separately." The third paper had a model answer and a grading rubric. The rubric allowed students to see that the majority of points were allocated to factual application and analysis.

Students were given about ten minutes of class time to use the model answer to do a self- or peer-edit.[89] The students then discussed the peer-edits with the person who did the edits. Five to ten minutes of class time were devoted to questions about the exercises or model answers.[90] After the in-class review, students completed a short anonymous questionnaire that sought information about the usefulness of the writing exercise and the review process.[91]

The intervention professor also graded and commented on one paper for each student. All students got their papers back after spring break.[92] Students who received only four points (below what was expected) were encouraged to meet with the intervention professor and other students were also given the opportunity to meet one-

---

89. Two papers involved self-assessment; one involved a peer edit.

90. The students also were free to ask questions on the class e-mail discussion list, although none of them did this.

91. For a brief discussion of student feedback, see *infra* Part V.E.

92. In order to spread her workload out throughout the semester, the intervention professor graded and commented upon approximately twenty-five papers per assignment. She used a numerical score of four points (below what was expected), six points (what was expected) or eight points (above what was expected). She also wrote comments on each paper. The professor found that grading different papers led to some unanticipated problems. For example, some students felt that they had been graded on "harder" papers and thus were disadvantaged. Others felt it was unfair that some students had the benefit of earlier model answers and thus had an advantage for later papers. In retrospect, given the various complexities of the different questions, both these student concerns had some merit. Additionally, some paper topics provided more fertile ground for meaningful feedback, while others were more *sui generis*. Thus, in future years, to the extent papers are graded, the same paper will be graded for all students and it will be one in which the feedback is most easily transferable. However, to keep students motivated, students will not be told which paper will be graded. Given how much time the grading took, it would be interesting to repeat this study without the individual grading component to see if the individual grading made any difference in outcome.

on-one with the intervention professor to discuss their papers. Few students took advantage of this opportunity. After this process was complete, about ten weeks into the semester, the students had a final essay question assignment in which they had the opportunity to incorporate the individual feedback.[93] There was also a model answer, peer edits, and in-class discussion for that final paper.

During the semester, the intervention and nonintervention professors decided to jointly test on Federal Rule of Civil Procedure 15 (amendments to complaints) and either claim or issue preclusion. For these substantive areas, they collaborated to ensure they were using the same cases and materials and were teaching the material from a similar perspective and with emphasis on similar key points. The professors waited until classes had ended before working together to draft the essay exam questions and model answers. In this way, they hoped to avoid inadvertently slanting their teaching in a way that would affect students' essay answers.

Each exam essay question was worth one-third of the exam's total raw score points and had a suggested time allocation of one hour.[94] Each essay question contained about one page of facts and it identified the overall issue. The professors designed the essay questions to test students' ability to break a rule into its component parts, to recognize relevant facts and how those facts corresponded to a particular element of a rule, and to analyze and apply the facts to the applicable constituent element. These were the specific skills the intervention professor had hoped to teach through the practice writing exercises.

The professors jointly devised a grading rubric for each question. They allocated the vast majority of the points to factual analysis, with a few points allocated to identifying component parts of the applicable rule and three points allocated to writing style and organization. To retain anonymity, all of the 135 exam answers (eighty from the intervention class and fifty-five from the nonintervention class)[95] were combined by the administrative support staff and all identifying characteristics other than the anonymous exam numbers were re-

---

93. The last paper was a one-page factual scenario involving a summary judgment issue. The final paper was worth three raw score points—two points for turning it in, and one point if the student attended the class in which papers were exchanged and peer-edited. Thus, students who turned in all five papers could receive a maximum of twenty raw score points and a minimum of sixteen raw score points toward their final grade.

94. Each professor made up her own short answer questions that were worth the remaining third of the exam's raw score points.

95. Both professors had some part-time second-year law students in their respective classes. These students were in their fifth semester of law school. Because we did not have easy access to all the data on the part-time students, only first-year students were included in this study's analysis. Thus, the sample size was seventy students for the intervention class and fifty-one for the nonintervention class.

moved. Each professor then began to grade all 135 essay answers to each essay question.

At the start of grading each essay question, the professors met and graded the first twenty-five questions together to ensure they were applying the rubric in a similar manner. After that, they periodically conferred to ensure that they continued to apply the rubric as consistently as possible.[96] After grading was completed, the administrative support staff put together a spreadsheet that contained the students' exam numbers. The spreadsheet contained the following information: each professor's grades on each essay question for all 135 students in the two classes, each student's UGPA and LSAT score, and each student's fall and spring semester grades in their other law school courses.

Using the information on the spreadsheets, we[97] sought to answer the following questions: 1) could we design a cost-effective duplicable model for assessing the impact of particular teaching methodologies on a specific learning outcome; 2) using the empirical model we designed, could we determine if the writing interventions produced a statistically significant difference in essay raw score points on each question between the students who received the writing interventions and those who did not; and 3) could we measure whether the interventions had any spillover effects on students' grades in other spring semester classes?

IV. METHODOLOGY

Our first and most important goal was to design a cost-effective empirical model for measuring the impact of a particular teaching innovation. Specifically, we wanted to measure whether a particular set of writing interventions could produce a statistically significant improvement on specific skills related to legal analysis. As discussed

---

96. The professors had a strong degree of concordance in their exam assessment. Out of a total possible thirty raw score points on Essay One, in only ten of the exams was there a variance of greater than four points. Out of a total of thirty-three possible raw score points on Essay Two, the professors had more than a four point variance on only fifteen exams. In cases where the point spread was more than four points, both professors regraded the exam in question. If it was determined that an exam warranted a different grade, the professor changed the original raw score grade. The authors of this Article believe the strong concordance was a result of the methodology used for grading—grading the first twenty-five together and discussing how/why the professors allocated points the way that they did after each batch of five exams were graded, and then periodically conferring throughout the grading process.

97. In Part III, Study Design and Implementation, we have referred to the intervention professor and nonintervention professor in the third person to describe how those coauthors were involved in the classroom teaching and grading. From this point forward, as was true in the previous sections other than Part III, the terms "we" and "our" refer to all three coauthors.

herein, we have created a model that can be duplicated and improved upon by others.

Our second goal was to see if the writing interventions had a discernible impact on students' ability to dissect a legal rule and perform a complex factual analysis during a final exam. We began this inquiry by looking at the raw scores on two essay questions, blindly graded by both professors—comparing those students in the writing intervention class with those in the nonintervention class. The results of independent sample t-tests[98] showed that, on average, there was a difference in performance on both essays, for both graders, with those students receiving the intervention fairing better in all four cases.

TABLE 1: AVERAGE RAW POINT ESSAY SCORES
(out of a total of 30.00 possible raw score points for Essay One and 33.00 raw score points for Essay Two)

	AVG SCORE ESSAY ONE	AVG SCORE ESSAY TWO
IC CLASS GRADED BY ICP	19.92	14.86
IC CLASS GRADED BY NICP	19.19	15.32
NIC CLASS GRADED BY ICP	17.27	11.98
NIC CLASS GRADED BY NICP	17.09	12.29

IC = Intervention Class; NIC = Nonintervention Class;
ICP = Intervention Class Professor; NICP = Nonintervention Class Professor

Clearly, this analysis alone is subject to substantial criticism for lack of control over a myriad of other factors besides the interventions that could vary between the two classes, including, for example, the difference in the professors teaching experience,[99] teaching style and possible stratification of student ability.[100] To address these concerns as much as is possible given the fact that the main goal was to teach Civil Procedure, and not to design an unassailable scientific experiment, we first examined whether there was some unidentified sorting mechanism by which more students with higher law school

---

98.	An independent sample t-test compares the mean scores of two groups on a given variable. The p-value states the likelihood that we would find a difference in means as large as that found purely by chance. A low p-value indicates that the difference is not likely due to chance. *See generally* COLLIN J. WATSON ET AL., STATISTICS FOR MANAGEMENT AND ECONOMICS 406 (5th ed. 1993).

99.	The intervention professor had twelve years of teaching experience; the nonintervention professor had been teaching for five years at the time this study occurred.

100.	Of course, other factors may also have impacted the results, such as students' other substantive course professors, students' undergraduate training, or any number of factors we may not have identified and isolated. We again note that given that the primary goal was to teach Civil Procedure, we make no claim that the results are unimpeachable. However, we do suggest that the controls we describe in this section helped give us confidence in our results and are controls that others seeking to duplicate this work can use or expand upon.

grade predictors[101] ended up in the intervention class. Using LSAT scores and UGPAs as indicators of predicted law school performance, an independent sample t-test indicated no significant difference between the students in the two classes on either of the measures (LSAT: p = .665, UGPA: p = .204).[102] At the same time, our analysis confirmed that both LSAT scores and UGPA were good predictors of overall performance in the law coursework offered that spring at the school where the study occurred.[103] These results increased our confidence that the differences in raw essay exam scores between the intervention and nonintervention classes were not the result of some unidentified sorting of students predicted to perform better on law school exams.

---

101.   The traditional data used to predict first-year law school grades is UGPA and LSAT score with the authors of the LSAT claiming that it is the better predictor. For a discussion of the correlation between these predictors and first-year law school performance, see *supra* notes 69-72. Although the use of LSAT scores for their predictive value has been criticized, see *supra* notes 70-71, given the prevalence of the use of these predictors, we used these variables to examine both whether more students predicted to achieve higher law school grades were in the intervention professor's class and to fine-tune our findings about the impact of the writing interventions themselves.

102.   *See* WATSON ET AL., *supra* note 98.

103.   *See infra* Table 2. In this table, we use a Pearson correlation. A Pearson correlation varies between -1 and 1, with a 1 indicating perfect correlation (the two measures vary together), a -1 indicating perfect negative correlation (when one measure increases the other decreases), and 0 indicating no correlation. The p value states the likelihood that we would find a correlation at this level due to chance. *See generally* WATSON ET AL.*, supra* note 98*,* at 926. We performed these correlations on the entire first-year class of 2005-06 at the subject school. We note that we found it interesting, and an area for future potential exploration, that while LSAT scores did a good job of predicting overall spring semester performance (correlation with the overall spring average = .286, p = .001), UGPA did a slightly better job (correlation with the overall spring average = .349, p = .000) with this particular cohort. *Id.* Further, we note that while there are some courses (Property and Legal Writing) where performance is poorly predicted by LSAT scores (Property: correlation = .128, p = .074; Legal Writing: correlation = .062, p = .386), UGPA is correlated with performance in all spring courses with significance levels p < 0.01. *Id.* As an aside, and as an area for future potential study, we note that amongst this study's cohort, LSAT scores are not correlated well with UGPAs (correlation = -.054, p = .427), suggesting that the two measures are not measuring the exact same thing. *Id.*

**TABLE 2: DO LSAT SCORES AND UNDERGRADUATE GPAs
PREDICT PERFORMANCE IN LAW SCHOOL?**

		LSAT	UGPA
LSAT	Pearson Correlation Sig. (2-tailed)		-.054 .427
UGPA	Pearson Correlation Sig. (2-tailed)	-.054 .427	
CIVIL PROCEDURE	Pearson Correlation Sig. (2-tailed)	.215* .011	.278** .001
CONTRACTS	Pearson Correlation Sig. (2-tailed)	.241** .001	.256** .001
TORTS	Pearson Correlation Sig. (2-tailed)	.240** .004	.278** .001
PROPERTY	Pearson Correlation Sig. (2-tailed)	.128 .074	.215** .003
LEGAL WRITING	Pearson Correlation Sig. (2-tailed)	.062 .386	.270** .000
CRIMINAL LAW	Pearson Correlation Sig. (2-tailed)	.286** .000	.232** .001
OVERALL SPRING AVERAGE	Pearson Correlation Sig. (2-tailed)	.286** .001	.349** .000

*Correlation significant at 0.05 level   **Correlation significant at 0.01 level

**TABLE 3: DO LSAT SCORES AND UNDERGRADUATE GPAs
PREDICT PERFORMANCE ON THE STUDY ESSAYS?**

		LSAT	UGPA
NIP'S GRADES FOR ESSAY ONE	Pearson Correlation Sig. (2-tailed)	.144 .119	.279** .002
NIP'S GRADES FOR ESSAY TWO	Pearson Correlation Sig. (2-tailed)	.120 .195	.269** .003
IP'S GRADES FOR ESSAY ONE	Pearson Correlation Sig. (2-tailed)	.155 .094	.303** .001
IP'S GRADES FOR ESSAY TWO	Pearson Correlation Sig. (2-tailed)	.089 .337	.256** .005

*Correlation significant at 0.05 level   **Correlation significant at 0.01 level
NIP = Nonintervention Professor; IP = Intervention Professor

Next, we stratified the data by LSAT scores and UGPA, dividing the data by the law school's overall median for that year's entering first-year class[104] (LSAT = 159, UGPA = 3.4), and conducted separate analyses. We began with each indicator separately. We discovered a number of interesting things we did not expect. First, we discovered that if one did not look at the intervention, UGPA was a strong predictor of performance on all four of the essay performance measures, while LSAT scores did not correlate with scores on the essay ques-

---

104. The "overall median" is that score that divides the high scoring half of the class from the low scoring half of the class. These medians were computed over the entire first-year class of 2005-06.

tions.[105] In other words, looking at all students without regard to which received the interventions, in general, students with high UGPAs received higher raw score essay question grades than those with lower UGPAs, while a student with a low LSAT score was just as likely to score well on the essay question as a student with a high LSAT score. However, once we added the interventions as a variable, we found that the interventions seemed to benefit those students with an above-the-median LSAT score most measurably. In fact, when looking at above-the-median LSAT scores in relation to the interventions variable, an independent sample t-test indicated highly statistically significant differences[106] in performance between intervention and nonintervention students on both essays across both graders.[107] A similar examination of intervention and nonintervention students with below-the-median LSAT scores revealed no such performance difference.[108] Thus, there was no statistically significant difference in raw essay scores between those below-the-median LSAT students who had the interventions and those below-the-median LSAT students who did not have the interventions.

TABLE 4: AVERAGE ESSAY SCORES FOR INTERVENTION AND NONINTERVENTION STUDENTS, BOTH WITH ABOVE-THE-MEDIAN LSAT SCORES

	AVG SCORE FOR STUDENTS IN INTERVENTION GROUP	AVG SCORE FOR STUDENTS IN NONINTERVENTION GROUP
NIP'S GRADES FOR ESSAY ONE	20.88	16.68
NIP'S GRADES FOR ESSAY TWO	16.08	12.37
IP'S GRADES FOR ESSAY ONE	21.33	16.90
IP'S GRADES FOR ESSAY TWO	15.66	11.75

Intervention Students = 36; Nonintervention Students = 32
(NIP1: p = .000; NIP2: p = .000; IP1: p= .000; IP2: p = .000)

---

105. *See supra* Table 3.
106. *See supra* note 103 and accompanying text; *see also supra* Table 1.
107. *See infra* Table 4.
108. *See infra* Table 5.

TABLE 5: AVERAGE ESSAY SCORES FOR INTERVENTION AND NONINTERVENTION
STUDENTS, BOTH WITH BELOW-THE-MEDIAN LSAT SCORES

	AVG SCORE FOR STUDENTS IN INTERVENTION GROUP	AVG SCORE FOR STUDENTS IN NONINTERVENTION GROUP
NIP'S GRADES FOR ESSAY ONE	17.22	17.78
NIP'S GRADES FOR ESSAY TWO	14.45	12.42
IP'S GRADES FOR ESSAY ONE	18.29	17.89
IP'S GRADES FOR ESSAY TWO	13.90	12.36

Intervention Students = 31; Nonintervention Students = 19
(NIP1: p = .679; NIP2: p = .123; IP1: p = .750; IP2: p = .264)

Examining only those students with above-the-median UGPAs, an independent sample t-test indicated highly statistically significant differences in performance on both essays across both graders between intervention and nonintervention students.[109]

TABLE 6: AVERAGE ESSAY SCORES FOR INTERVENTION AND NONINTERVENTION
STUDENTS, BOTH WITH ABOVE-THE-MEDIAN UGPA

	AVG SCORE FOR STUDENTS IN INTERVENTION GROUP	AVG SCORE FOR STUDENTS IN NONINTERVENTION GROUP
NIP'S GRADES FOR ESSAY ONE	20.67	17.9
NIP'S GRADES FOR ESSAY TWO	16.67	13.86
IP'S GRADES FOR ESSAY ONE	21.41	18.20
IP'S GRADES FOR ESSAY TWO	16.32	13.33

Intervention Students = 34; Nonintervention Students = 30
(NIP1: p = .005; NIP2: p = .005; IP1: p = .000; IP2: p = .009)

A similar examination of intervention and nonintervention students with below-the-median UGPAs revealed a statistically significant performance difference on the second essay question only, but across both graders.[110]

---

109. *See infra* Table 6.
110. *See infra* Table 7.

TABLE 7: AVERAGE ESSAY SCORES FOR INTERVENTION AND NONINTERVENTION STUDENTS, BOTH WITH BELOW-THE-MEDIAN UGPA

	AVG SCORE FOR STUDENTS IN INTERVENTION GROUP	AVG SCORE FOR STUDENTS IN NONINTERVENTION GROUP
NIP'S GRADES FOR ESSAY ONE	17.66	15.95
NIP'S GRADES FOR ESSAY TWO	13.94	10.28
IP'S GRADES FOR ESSAY ONE	18.39	15.95
IP'S GRADES FOR ESSAY TWO	13.36	10.04

Intervention Students = 33; Nonintervention Students = 21
(NIP1: p = .244; NIP2: p = .001; IP1: p = .058; IP2: p = .004)

To further isolate the effect of LSAT score and UGPA on our outcomes, we then conducted analyses using both indicators to separate the data into four strata: above-the-median LSAT and above-the-median UGPA; above-the-median LSAT and below-the-median UGPA; below-the-median LSAT and above-the-median UGPA; and below-the-median LSAT and below-the-median UGPA.[111]

Examining students with above-the-median LSAT scores and above-the-median UGPAs showed a statistically significant difference in performance on both essays across both graders between students who had the interventions and those who did not.[112]

TABLE 8: AVERAGE ESSAY SCORES FOR INTERVENTION AND NONINTERVENTION STUDENTS, BOTH WITH ABOVE-THE-MEDIAN LSAT SCORES AND ABOVE-THE-MEDIAN UGPA

	AVG SCORE FOR STUDENTS IN INTERVENTION GROUP	AVG SCORE FOR STUDENTS IN NONINTERVENTION GROUP
NIP'S GRADES FOR ESSAY ONE	21.57	17.7
NIP'S GRADES FOR ESSAY TWO	16.89	13.64
IP'S GRADES FOR ESSAY ONE	22.15	18.05
IP'S GRADES FOR ESSAY TWO	17.00	13.00

Intervention Students = 19; Nonintervention Students = 17
(NIP1: p = .005; NIP2: p = .013; IP1: p = .002; IP2: p = .006)

For students with above-the-median LSAT scores and below-the-median UGPAs, the differences in performance between intervention

---

111.   These sample sizes were necessarily smaller than our other sample sizes and thus we report these results with the caveat that the lack of statistical significance may be influenced by the small sample size.

112.   *See infra* Table 8.

and nonintervention students remained statistically significant across the board.[113]

TABLE 9: AVERAGE ESSAY SCORES FOR INTERVENTION AND NONINTERVENTION STUDENTS, BOTH WITH ABOVE-THE-MEDIAN LSAT SCORES AND BELOW-THE-MEDIAN UGPA

	AVG SCORE FOR STUDENTS IN INTERVENTION GROUP	AVG SCORE FOR STUDENTS IN NONINTERVENTION GROUP
NIP'S GRADES FOR ESSAY ONE	20.11	15.53
NIP'S GRADES FOR ESSAY TWO	15.17	10.93
IP'S GRADES FOR ESSAY ONE	20.41	15.60
IP'S GRADES FOR ESSAY TWO	14.17	10.33

Intervention Students = 17; Nonintervention Students = 15
(NIP1: p = .009; NIP2: p = .004; IP1: p = .001; IP2: p = .009)

For students with below-the-median LSAT scores and above-the-median UGPAs, no statistically significant difference in performance was found between intervention and nonintervention students.[114]

TABLE 10: AVERAGE ESSAY SCORES FOR INTERVENTION AND NONINTERVENTION STUDENTS, BOTH WITH BELOW-THE-MEDIAN LSAT SCORES AND ABOVE-THE-MEDIAN UGPA

	AVG SCORE FOR STUDENTS IN INTERVENTION GROUP	AVG SCORE FOR STUDENTS IN NONINTERVENTION GROUP
NIP'S GRADES FOR ESSAY ONE	19.53	18.15
NIP'S GRADES FOR ESSAY TWO	16.40	14.15
IP'S GRADES FOR ESSAY ONE	20.46	18.38
IP'S GRADES FOR ESSAY TWO	15.46	13.76

Intervention Students = 15; Nonintervention Students = 13
(NIP1: p = .310; NIP2: p = .167; IP1: p = .089; IP2: p = .372)

For the students where both LSAT scores and UGPAs were below-the-median, there was not a statistically significant difference in performance except in the case of the nonintervention professor's evaluation of the second essay.[115]

---

113. *See infra* Table 9.
114. *See infra* Table 10.
115. *See infra* Table 11.

TABLE 11: AVERAGE ESSAY SCORES FOR INTERVENTION AND NONINTERVENTION
STUDENTS, BOTH WITH BELOW-THE-MEDIAN LSAT SCORES AND BELOW-THE-
MEDIAN UGPA

	AVG SCORE FOR STUDENTS IN INTERVENTION GROUP	AVG SCORE FOR STUDENTS IN NONINTERVENTION GROUP
NIP'S GRADES FOR ESSAY ONE	15.06	17.00
NIP'S GRADES FOR ESSAY TWO	12.62	8.66
IP'S GRADES FOR ESSAY ONE	16.25	16.83
IP'S GRADES FOR ESSAY TWO	12.5	9.33

Intervention Students = 16; Nonintervention Students = 6
(NIP1: $p = .451$; NIP2: $p = .033$; IP1: $p = .803$; IP2: $p = .106$)

While these results did not eliminate the possibility that the differences in performance found in this study may be attributable to differences between the two classes other than the writing interventions, they increased our confidence that the interventions were responsible for the differences for two reasons. First, if one professor had even inadvertently "taught to the test," we would have expected to see statistically significant differences in performance between the intervention and nonintervention students across all strata. That is to say that we would have expected all students who had the writing intervention, regardless of LSAT score, to do measurably better than the students who did not have the intervention.[116] The same would likely be true if the findings were due to the different teaching styles or experience levels of the two professors. Second, as discussed below, there is theoretical consistency in our findings that students with above-the median LSAT scores and above-the-median UGPAs were better able to generalize the lessons offered by the writing interventions and apply them to novel problems than students with below-the-median LSAT scores and, to some extent, below-the-median UGPAs.[117]

The last question we examined was whether these performance gains on the anonymous, jointly graded essays translated to better grades in other courses. We approached this question in two ways. First, we examined the difference in final exam grades in other substantive first-year classes for all students in both the intervention

---

116. *See supra* Tables 4-5 and accompanying text. Although there was a raw score point difference in both essay questions with the Intervention Professor's students, on average, getting higher raw score points on both essays (see *supra* Table 1) that point spread did not rise to the level of statistical significance. However, as noted in Tables 4-5, there was a statistically significant difference in performance between intervention and nonintervention students with above-the-median LSAT scores, but no such statistically significant difference in students with below-the-median LSAT scores.

117. *See* discussion *infra* Parts V.B-C.

section and the nonintervention section. For example, examining grades in Professor X's Torts class, we compared intervention class students in Professor X's Torts class with the nonintervention class students in Professor X's Torts class. An independent sample t-test revealed no significant differences in final exam grades in any of the students' other substantive courses.

We also examined performance in other classes accounting for LSAT and UGPA. For example, we compared intervention class above-the-median LSAT score students in Professor X's Torts class to nonintervention class above-the-median LSAT score students in Professor X's Torts class. Again, an independent sample t-test revealed no significant differences in performance except for one section of a first-year Torts class, where the students who did not have the intervention performed better on that entirely multiple choice final exam. Possible explanations for the findings that the interventions had no "spillover" effect in other courses are discussed in more detail in Part V.D.

V. DISCUSSION OF FINDINGS

A. *Writing Interventions Had an Impact on Raw Essay Scores*

As discussed in the previous Part, we found that, on average, students who received the writing interventions had higher average raw scores on each of the essay questions.[118] Although this point spread did not rise to the level of statistical significance, we found between a two and three point raw score difference (out of a possible thirty raw score points on the first question and thirty-three raw score points on the second question).[119] These findings are somewhat consistent with the literature that indicates that, for undergraduates, practice tests contribute to improved performance on actual tests,[120] especially to the extent the practice exams closely match the format and difficulty of the actual exam.[121] However, most of those studies did not stratify students by SAT scores. With regard to the one study that did stratify students by SAT score, our results varied from that study's findings that improvement occurs across the board, regardless of SAT scores.[122] As noted herein, we found that, on average, the interventions improved the scores of students with above-the-median LSATs, but did not have any statistical impact on stu-

---

118. *See supra* Table 1.
119. In practical terms, if the same students had a raw point score difference of three points on each question, those six raw score points could mean about a half-step difference in their final grade (e.g., their grade would go from a seventy-seven to an eighty, or an eighty to an eighty-three).
120. *See supra* notes 33-40 and accompanying text.
121. *See supra* notes 34-35 and accompanying text.
122. *See supra* notes 39-40 and accompanying text.

dents with below-the-median LSAT scores.[123] This difference is perhaps explainable by the fact that the study using SAT scores as a variable dealt with undergraduate exams involving multiple choice questions, the same type of questions used on the SAT. In our study, the essay questions obviously vary in format from LSAT multiple-choice questions. Thus, this study explored a new question: do practice essay tests have the same benefit for all students regardless of LSAT or UGPA? Our initial results indicate that, in fact, the benefit of practice essay tests does not inure equally to all students. In the next subpart of this Article, we explore potential explanations for our finding that those students with above-the-median LSAT scores and, to some extent, above-the-median UGPAs seemed to have derived the most measurable benefit from the practice writing exercises.

B. Why Did the Interventions Have the Most Statistically Significant Benefit for Students with Above-the-Median LSAT Scores and Above-the-Median UGPAs?

As discussed above, to help control for variables such as teaching style, teaching experience, and students' general test-taking abilities, we examined the relationship of LSAT scores and UGPA to the raw essay scores. In doing so, we made two relatively surprising discoveries. First, UGPA was a strong predictor of students' raw essay scores even if we did not consider the intervention. Thus, students with above-the-median UGPAs, on average, scored higher than students with below-the-median UGPAs regardless of whether they were in the intervention or nonintervention class.[124] When we factored in the intervention, we found a statistically significant difference in performance between intervention and nonintervention students with above-the-median UGPAs.[125] This difference existed as to both essay questions across both graders. Between intervention and nonintervention students with below-the-median UGPAs, we found a statisti-

---

123.   *See supra* Tables 4-5.
124.   *See supra* Table 3. On average, there was a statistically significant two-point raw score difference between the above-the-median and below-the-median UGPA students for each essay across each grader. *Id.* The finding that there was a statistically significant difference based upon UGPA regardless of the intervention is consistent with our finding that UGPA correlated to Research, Writing, and Advocacy (RWA) grades while LSAT scores had no correlation with RWA grades. *See supra* Table 2. This finding intuitively makes sense given that the LSAT, a multiple-choice test, does not reflect a student's ability to engage in actual legal writing and analysis while UGPA often involves grades for courses in which writing ability and written analysis play a role in the final grade. It is interesting to note that the two point spread was statistically significant when looking at the correlation between UGPA and raw essay scores and LSAT score and raw essay scores, but not statistically significant when looking at the correlation between the intervention and nonintervention classes as a whole. One possible explanation for this result is the different sample sizes in these two different cohort groups.
125.   *See supra* Tables 6-7.

cally significant performance difference on the second question for both graders.[126] Thus, the interventions seemed to be the most helpful to the intervention class students with above-the-median UGPAs, but were at least somewhat helpful to the intervention class students who had below-the-median UGPAs.[127]

On the other hand, LSAT score alone had no relationship to students' raw point essay scores.[128] Thus, if we did not factor in the intervention, students with above-the-median LSAT scores were no more likely to get high raw scores on the questions than students with below-the-median LSAT scores.[129] However, as noted earlier, for the 2005-06 first-year class of students at the school where this study occurred, LSAT score, unlike UGPA, had virtually no correlation to students' legal research, writing, and advocacy grades.[130] One potential explanation for our finding that there was generally no correlation between LSAT score and raw essay score until we factored in the

126. *See supra* Table 7. One reason for the difference in performance on the first and second essay questions may have been in the nature of the questions. Essay One, a claim preclusion issue, involved a more complex legal analysis and a less straightforward rule than Essay Two, which was a Rule 15 amendment and relation-back question. The relation-back analysis for Essay Two was a "same transaction or occurrence analysis"—something the intervention students had done in both the Rule 20 practice essay and the Rule 13(a) practice essay. The amendment analysis in Essay Two involved easily identifiable factors (e.g., whether there was undue delay, unfair prejudice, etc.) and the application of facts to these Rule 15 factors was more straightforward than the application of facts in Essay One, which involved more theoretical issues such as whether parties were in privity and whether virtual representation existed. In sum, although we did not think so when we wrote the questions, Essay One was analytically and factually more difficult than Essay Two, which involved, at least in part, an analysis of an issue the intervention students had analyzed in two practice problems. Thus, to the extent practice helped the intervention below-the-median UGPA students, it helped them the most when the essay question was more straightforward and more similar to the practice questions. This may also explain the results of Table 11, which shows that as to the below-the-median UGPA and below-the-median LSAT score students, there was a statistically significant difference in performance between the intervention and nonintervention professor as to Essay Two as graded by the nonintervention professor. *See supra* Table 11.

127. *See supra* Tables 6-7. The difference in the difficulty level of the two questions may also explain why we saw that the interventions had an impact on students with below-the-median UGPAs only with the second question. *See supra* note 126. Since the second question was more straightforward, and more similar to the practice questions, it may have been easier for all of the intervention professor's students to apply what they had learned through the writing interventions. However, the difficulty of the first question may have presented more difficulty in transferring that learning for students with lower level metacognitive skills. For a discussion of the interplay between metacognitive skills and the study's results, see *infra* Part V.C.

128. *See supra* Table 3. This is different from what was found by Gretes and Green when they examined the correlation between SAT scores, practice tests, and performance on actual tests. *See supra* notes 39-40 and accompanying text. The difference may be explained because the SAT was more similar in format to the computerized practice tests and the actual multiple-choice tests given in the Gretes and Green study. In this case, there is no similarity per se between the LSAT multiple-choice format and the essay final examination questions.

129. *See supra* Table 3.

130. *See supra* Table 2.

intervention is that although the LSAT may have some predictive value in terms of certain skills relevant to success in law school, the skills tested by these essay questions, as well as the skills measured by first-year legal research, writing, and advocacy (RWA) courses are not skills measured by the LSAT.[131]

After we saw the study results, one question we asked ourselves is why the interventions provided the most statistically significant benefit to students with above-the-median LSATs and above-the-median UGPAs. One explanation may be that the interventions tapped into students' metacognitive abilities. The literature suggests that students with more highly developed metacognitive skills—the ability to know when, why, and how to use a learning strategy—generally get higher undergraduate grades.[132] Additionally, given the existing literature on the relationship between metacognitive skills and reading comprehension,[133] and the ability to determine where to focus study efforts,[134] we wonder if there is a link between highly developed metacognitive skills and LSAT performance. Although we found no studies linking metacognitive skills to LSAT scores, given that the LSAT involves reading comprehension, practice test-taking, and learning how to identify and fix test-taking weaknesses, it is not unreasonable to hypothesize that students with more highly developed metacognitive skills are, on average, likely to get higher LSAT scores. Given that we know metacognitive skills correlate to UGPA,[135] and given the reasonable—but untested—hypothesis that metacognitive skills also correlate with LSAT scores, we think one explanation for our findings may be that students with more highly developed metacognitive skills were better able to make use of the information provided to them via the interventions. We explore this idea in the next subpart.

---

131. We think that the correlation between LSAT score and RWA grades merits further study. Given that RWA courses involve the application of skills in ways that practicing lawyers use these skills, if future studies indicate that LSAT scores fail to correlate to students RWA grades, law schools may need to rethink the use of the LSAT as an admissions criteria.

132. *See* Baker, *supra* note 45, at 9-10; Minnaert, *supra* note 54, at 238-39; Nietfeld et al., *supra* note 47, at 23.

133. *See supra* notes 48-53 and accompanying text (discussing the link between metacognitive skills and reading comprehension).

134. *See supra* notes 57-58 and accompanying text.

135. *See supra* notes 52-56; *see also* Schwartz, *supra* note 15, at 472-83.

C. How Metacognition May Have Played a Role in How Students Used the Interventions

1. Figuring Out What Would Be Tested

One aspect of highly developed metacognitive skills is the ability to figure out what is likely to be tested.[136] As Professor Kissam has noted, one of the hidden things we test is law students' ability to react to surprise.[137] As one study found, the best specific metacognitive strategy for getting a high test grade is verbalizing what the test-taker predicts the test question will be.[138] In Civil Procedure, there is a limited universe of testable subject matter. The intervention professor had already given essay questions in five of the main substantive areas covered during the semester and had also provided students with model answers and, in one instance, a grading rubric. Students with more highly developed metacognitive skills may have looked at what had been covered already, and predicted the major areas that had not been tested and which of the available subject areas would provide appropriate contexts for a one-hour exam question. Accordingly, they then may have focused their studies on those few remaining substantive areas that would lend themselves to a one-hour essay exam question. In fact, one high-performing student confirmed that she and her study group had done just that.[139] On the other hand, the nonintervention professor's students, who had no idea what was likely to be tested, had to be more comprehensive in how they studied and where they focused their study efforts as they prepared for the final exam. Certainly, this is a factor that should be considered by others replicating this study. It could be controlled by simply telling students in both classes the potential substantive areas covered by the essay questions while also telling them that all substantive areas would be fair game for the short answer questions.

2. Identifying and Improving Weaknesses

Another way metacognitive skills may have come into play is that those with more highly developed metacognitive abilities are better able to identify their problem areas and develop strategies for improvement.[140] Thus, students with more highly developed metacognitive abilities may have been better able to use the writing exer-

---

136.   *See supra* text accompanying notes 56-58.
137.   Kissam, *supra* note 22, at 453-54.
138.   Baker, *supra* note 45, at 9-10.
139.   One of the intervention professor's high performing students came by to review her exam. After the review, she volunteered that she and her study group had figured that it was likely that there would be a Rule 15 question and a claim or issue preclusion essay question because of what the intervention professor had covered in the practice questions.
140.   *See supra* notes 54-57 and accompanying text.

cises and feedback from the model answers, peer edits, self-edits, grading rubrics, and individualized feedback to identify their weaknesses in terms of analysis and application. Once they were able to identify their weaknesses, they also may have been able to use the information provided to figure out how to improve in those specific areas. For example, a student could have figured out that he or she needed to pay more attention to breaking a rule into its constituent parts. Or the student might have found that he or she better needed to apply the facts generally, or to each element, or to both sides of an issue. Whatever the weakness, the student with more highly developed metacognitive skills could have used the exercises and concomitant feedback to identify both his or her weakness and the solution to strengthening that weakness.

3. *Understanding and Transferring the Analytical Formula*

Another possible benefit that students with more advanced metacognitive abilities may have derived from the writing interventions is that the practice assignments afforded them the opportunity to discern and practice using the analytical formula required to do well on the final exam. Many professors teach this framework as IRAC (Issue, Rule, Analysis, and Conclusion) and encourage students to use it as a means of analyzing the legal problem presented, applying the controlling legal doctrine to the relevant facts, and organizing their response to the call of the question.[141] Knowledge of the formula and mastery of the intricacies inherent to its application are not synonymous competencies. Mastery of the skill requires repeated exposure to the formula and repetitious application of the formula in the context of different exam questions.[142]

The intervention professor's students had five opportunities to study, practice, and master the application of IRAC to a myriad of question contexts. Their exposure to several different types of questions presenting different legal principles and different factual scenarios,[143] as well as the feedback given on those different questions, afforded the opportunity to respond to the challenge of applying the analytical construct embodied in the IRAC formula in multiple contexts. Students with more highly developed metacognitive abilities may have been more adept at accurately assessing and correcting

---

141. The call of the question refers to the specific question that the students were asked to focus on and answer based on the facts and the applicable doctrine. Essentially, the call of the question limits the students' focus to the specific issue(s) being evaluated.

142. Schraw & Brooks, *supra* note 44, at 5 (referencing this skill as a "strategy," and encouraging teachers to "teach specifically for transfer by using the strategy in a variety of settings," explaining "the more automatic a strategy, the more likely it is to transfer").

143. *See supra* notes 85-86, 93 (describing the subject matter of the various practice essay questions).

any weaknesses in their use of the formula. By virtue of the repeated practice in applying IRAC and the opportunity to learn from the feedback, their exam answers would reflect a greater command of the application of that skill set (i.e., more effective application of the elements of each legal principle, richer factual analysis of the tested issues, and better organization of their answer) and a stronger ability to transfer the skills to different fact patterns and questions.[144]

Intervention students with less developed metacognitive skills may have understood the analytical formula but may not have been able to generalize its applicability to different factual and legal contexts. Essentially, they may not have been able to apply or transfer that framework to the factual and legal contexts presented by the final exam questions.

4. Improving Test-Taking Speed

In addition to metacognitive skills playing a role in students' ability to become more adept at transferring and applying the analytical framework, these skills may have intersected with another skill—test-taking speed (i.e., the speed at which one is able to analyze and respond to a test question) in a way that enhanced the students' performance on the two exam questions.[145] In a study testing the relationship between LSAT performance, law school exam performance, and test-taking speed, Professor Henderson reports that "test-taking speed is a variable that affects the ordinal ranking of students on both the LSAT and actual law school exams."[146] After determining that speededness is a variable common to both the LSAT and law school exams,[147] Professor Henderson concludes that "[a] student with a fast rate of test-taking speed will likely do better on the LSAT than a student with the same level of reasoning ability but a lower rate of test-taking speed"[148] and "students with fast test-

---

144. We suggest than an area for further study is the connection between metacognitive ability and LSAT performance. The metacognitive skills of figuring out what is expected on a test, learning to meet that expectation, and learning the method of analysis that will achieve the correct response is part of what is involved in practicing and preparing for the LSAT. Our own experience informs us that students who do well on the LSAT have often figured out the "rules of the game" in terms of the kinds of questions that will be asked and how to respond successfully to them. If this is true, it is not surprising that the same students who "learn the game" in order to crack the LSAT code are able to "learn the game" in order to crack the law school essay examination code. Additional research confirming this may provide another explanation for why the writing interventions and accompanying feedback in our study provided a greater benefit to above-the-median LSAT performers than they did to below-the-median LSAT performers.

145. *See supra* note 73 and accompanying text.

146. Henderson, *supra* note 19, at 1044.

147. *Id.* at 979.

148. *Id.* at 1044.

taking speed are clearly favored on in-class exams."[149] Just as students who repeatedly practice LSAT questions can improve their performance by being able to analyze and respond to the questions more quickly, the intervention professor's students with more highly developed metacognitive skills may have been able to become more adept at quickly transferring and applying the requisite analytical framework in their responses to the two timed Civil Procedure exam questions. Thus, it is possible that students with highly developed metacognitive skills who were able to engage in factual and legal analysis in a myriad of contexts with progressive efficacy developed greater competency in applying the formula. The competency allowed them to apply the formula more quickly. The students with less developed metacognitive skills who were unable to transfer the applicable analytical framework from the practice writing assignments may not have had the advantage of speedy application of that framework. Thus, students' metacognitive skills may have played a role in their ability to discover their general analytical strengths and weaknesses, figure out how to improve their weaknesses, understand and utilize the general analytical formula that was expected on the exam, and transfer what they had learned about the analytical formula to different legal issues and factual scenarios quickly.

D.   *Writing Interventions Did Not Affect Grades in Other Courses*

Another question we looked at was whether the writing interventions had any spillover effect; did students who had the interventions perform better on their final examinations in other substantive classes than students who did not have the interventions? We tried to measure whether this occurred by comparing the grades of intervention and nonintervention students in each of the students' other spring semester courses. We were able to control for two variables. First, we controlled for UGPA and LSAT score. Second, we controlled for variances in other professors' exam formats and grade normalization by only comparing writing intervention/nonintervention students in the same sections of other substantive classes.[150] We found

---

149.   *Id.* at 982.

150.   At the school where the study occurred, there are three sections of each substantive first-year required course: Civil Procedure, Contracts, Criminal Law, Property, and Torts. We compared intervention to nonintervention students within the same section against each other. For example, Professor X teaches one section of Torts. We compared the intervention and nonintervention students enrolled in Professor X's Torts section. We did the same thing for each of the students' spring semester substantive courses as well as their research and writing course. Necessarily, the sample size for this portion of the study was relatively small, ranging from a high of twenty-four intervention and twenty-four nonintervention students in one contracts section to a low of twenty intervention and thirteen nonintervention students in one Torts section. This small sample size could also be a reason for the lack of any statistically significant differences in grades between the intervention and nonintervention students in other courses.

the writing interventions had no discernible effect on students' grades in other spring semester substantive courses.

Although we initially were surprised and disappointed by this finding, in retrospect, there are a number of reasons that can explain why there was no "spillover" effect. First, the type of final exam varied widely between faculty members. For example, some professors gave mostly short answer questions or a mixture of short answer and multiple-choice questions and one professor gave solely a multiple-choice exam. The skills needed to do well on those types of exams were not skills taught by the writing interventions.[151] Second, the other final exams may have tested a much wider range of skills than those taught by the writing interventions. For example, identifying issues was not a component of the writing interventions and yet it may have played a significant role in a first-year final exam. Or a professor may have emphasized policy analysis and included that as a significant component of what was measured. Or a professor may have given a take-home exam, thus eliminating the speededness component.[152] Finally, to the extent that all exams tested a student's ability to dissect a legal rule and perform a complex factual analysis, students may not have been able to transfer the skills they learned in the intervention professor's Civil Procedure class to a different class. This would be consistent with the literature that suggests that "[n]ewly acquired strategies do not readily transfer to new tasks or unfamiliar domains,"[153] and that to impact student performance in courses other than the ones in which the skills are explicitly being used, one must spend a significant amount of time helping students acquire an understanding of the metacognitive skills necessary to understand how to transfer the skills to other courses.[154] Or, perhaps, it is simply that students with more highly developed metacognitive

---

151. *See* Wilbert J. McKeachie et al., *Teaching Learning Strategies*, 20 EDUC. PSYCHOLOGIST 153, 158 (1985) (noting that "[l]earning strategy training programs must be sensitive to the match between learning activities and the criterial tasks used to assess performance" and that different strategies may be appropriate for essay exam, multiple-choice exam, or research paper assessments).

152. In fact, these three variables came into play in the three Torts sections. One Torts professor gave a wholly multiple-choice exam, the second gave a very policy-oriented exam, and the third gave a take-home exam.

153. Schraw & Brooks, *supra* note 44, at 5. Professors Schraw and Brooks note that research indicates that if one teaches strategy instruction, one should make sure to teach students how to transfer the strategy by having the students use that strategy in a number of different situations. *Id.*

154. *See generally* Masui & De Corte, *supra* note 59 (describing a comprehensive integration of metacognitive skills training into a substantive college course); Niedwiecki, *supra* note 42 (arguing that law professors should help students learn metacognitive skills in substantive courses); Schwartz, *supra* note 15, at 484-505 (describing a class in building metacognitive skills and discussing how to integrate metacognitive skill-building into regular first-year law courses).

skills predicted what would be tested in Civil Procedure and could not make those same predictions for other classes.[155]

E. *Student Response to the Writing Interventions*

Throughout the semester, the intervention professor asked students for anonymous feedback about the usefulness of the exercises and the different forms of feedback (model answers, peer-edits, and self-edits). The students were given a Likert scale[156] to rate the usefulness of the exercises to their understanding of: the substantive law, how to approach an essay question, and their own strengths and weaknesses. They also were asked for general comments. At the end of the semester, the students were asked if the interventions should be given in future years.

For all four exercises in which student feedback was solicited, the overwhelming majority of the students rated the exercises as helpful to very helpful to their learning of the substantive material. The vast majority of students also felt the exercises demystified the exam-writing process and helped them better understand how to analyze legal issues. In fact, despite the extra work involved in writing five short papers throughout the semester, seventy-one out of seventy-two students said that the exercises should be given to students in future classes.

When asked about the specific forms of feedback, virtually all students appreciated getting the annotated model answers and felt that these helped them better understand what was expected.[157] Student reaction about the usefulness of the rubric was more mixed, with some students finding it helpful and some finding it confusing. Finally, most students also found the self-editing and peer-editing process helpful. One student said, "definitely continue the self-review and/or classmate review. I am seeing edits to my work that point out problems I have in all my law school essay work. Thank you for this

---

155. *See supra* text accompanying notes 57-58 (discussing the role of predicting what will be on an exam as an important skill and how that skill could have been used in the context of this study). We also acknowledge that there may be other explanations for the lack of "spillover" effect that we have not even considered.

156. A Likert scale is a scale frequently used in questionnaires in which respondents specify a level of response to a question (i.e. very helpful, helpful, not helpful, etc.). *See* Rensis Likert, *A Technique for the Measurement of Attitudes*, 140 Archives Psychol. 5 (1932). For a discussion of how to use Likert scales, see generally EARL R. BABBIE, THE BASICS OF SOCIAL RESEARCH 174 (3d ed. 2005).

157. Interestingly, given that the interventions did not seem to help students with below-the-median LSAT scores, this self-assessment may bolster the argument that metacognitive skills played a role in the effectiveness of the interventions. As noted earlier, often students with poor metacognitive skills do not understand their own comprehension weaknesses. *See supra* note 57. Thus, students with less developed metacognitive skills may have thought the exercises were helpful when, in fact, the exercises and feedback did not actually help them identify and remedy their weaknesses.

help!" Some students felt the individualized feedback was helpful while others felt that they could not transfer the feedback to a different fact pattern.[158] Many students would have liked more individualized feedback. As one student noted:

> The only problem I had was that I wish I could've gotten feedback on my other papers especially the later ones. I know you don't have time and that it's an unreasonable request, but I felt like I improved and I wanted some feedback to see if I really was making progress—not your fault at all, just wishful thinking.

Finally, when asked how these exercises could be more helpful, a number of students suggested that in future years, they should get some essay questions with multiple issues so that the essay questions would more closely mirror what they had experienced in their first-semester law school exams.

In sum, despite the extra work involved for the students, almost all the students felt that the exercises and accompanying feedback were useful. Virtually all students thought that the exercises should be continued in future years, although many students suggested that it would be more helpful if the exercises were accompanied by more individualized feedback. Given our findings that the interventions did not provide across-the-board, statistically significant benefits to all students, it is interesting that the intervention students' responses were generally so uniformly positive. These responses reinforce what both the Best Practices and Carnegie Foundation studies have emphasized: law students want, and need, meaningful formative assessment opportunities.[159] Given the lack of formative assessment opportunities in other courses, the positive student responses to these exercises and feedback may indicate that students simply appreciated a chance to apply and practice some skills and receive even generalized feedback on their performance. We believe that the students' responses to these exercises certainly support the call for the use of more formative assessments throughout the semester.

VI.   FUTURE AREAS FOR FURTHER STUDY

We do not presume to claim that this study definitively indicates that writing interventions should be incorporated into first-year substantive classes. We think it is up to individual professors to decide if they believe this study's results support the additional time investment required to give students take-home essay questions with feedback. However, we do think that the study lays the groundwork for

---

158.   We agree with this assessment. Some of the practice questions lent themselves to feedback that could be better generalized.

159.   STUCKEY ET AL., *supra* note 1, at 255-59; SULLIVAN ET AL., *supra* note 1, at 164-73.

further investigation regarding the value of writing exercises and other teaching innovations.

There are many areas ripe for further study. For example, we wonder if the writing interventions would have been of greater benefit to students with below-the-median LSAT scores if we had more explicitly taught metacognitive skills. Would we have gotten different results if the intervention professor had incorporated reflective exercises designed specifically to help students evaluate their strengths and weaknesses and to help them develop specific steps to deal with their weaknesses?[160]

Strategies that could be incorporated into classes using writing interventions or other teaching innovations include things such as asking students some of the following questions: "Where did you look to determine the rule's constituent elements?"; "What did you do to clarify your understanding of a confusing element before you began writing?"; "How did you decide which facts to use to support your argument?"; "Why did you choose those facts?"; "Did you use all the relevant facts?"; "What will you repeat and what will you change in how you go about identifying the elements of a rule and/or choosing which facts to apply to your arguments?"

Also, after giving students an opportunity to review an annotated model answer, professors could ask students to make a list of their mistakes and a list of things they did well. Students could be asked to identify five steps they would follow in approaching future essay questions. After the next essay question was completed, they could be asked which steps proved to be useful strategies and which were not that useful. Additionally, the class, as a whole, could discuss useful strategies and study methods such as: do a decision tree before writing an answer; don't forget to check the plausibility of an argument; always use the word "because" and follow it with an explanation; and double-check to make sure you have considered both sides of an argument.[161] It would be interesting to examine whether this kind of active reflection results in the interventions having a statistically significant benefit for all students and to see if this kind of metacognitive skills instruction helps students more easily trans-

---

160. *See generally* Masui & De Corte, *supra* note 59 (discussing their findings that reflecting upon which learning strategies, learning aids, and allocation of time and effort contributed to learning outcomes is a key metacognitive activity that can be taught and can have an impact on student performance in multiple courses).

161. To motivate students to take these metacognitive strategy sessions seriously, the professor should make sure to explain why students are doing the reflective exercises because others have found that this motivates students to engage actively in the metacognitive exercises. *See* Schraw & Brooks, *supra* note 44, at 14-16 (discussing the connection between motivation and using metacognitive strategies and the necessity of discussing and explaining the value of metacognitive strategies prior to teaching students the strategies).

fer what they have learned in one class to their other classes and exams.[162]

Another question we have is whether our results would have been different if we had given a take-home, rather than an in-class timed exam. This change would eliminate the speededness component that impacts both LSAT score and in-class timed exams.[163] In addition, this change would make the exams more similar to the practice exams, something that others have found increases the impact of practice exercises on final exam performance.[164] It would be interesting to see if giving a take-home exam results in a statistically significant benefit to all intervention class students rather than concentrating the measurable benefit on those with above-the-median LSAT scores and above-the-median UGPAs.[165] Certainly, these questions are ripe for further study.

We also think it would be interesting to isolate the factors we examined and to see if the same results would occur if the intervention professor used only one form of generalized feedback such as only annotated answers, or annotated answers only with peer edits. We also wonder if the interventions would have had a more significant impact on below-the-median LSAT score students if, rather than becoming increasingly more difficult, they continued to focus on very basic skill set acquisition, perhaps by giving a writing exercise with generalized feedback on a relatively straightforward rule, and then, after the annotated answer, giving a second exercise using a different factual scenario applying that same rule.

Finally, we recognize that there are some variables that we did not control that could have affected the study's outcome. For example, we think that in order to eliminate the variable of the impact of teaching experience and style, future studies would be more reliable if the same professor taught two different sections of the same substantive course. Likewise, we cannot eliminate the possibility that other factors, such as the impact of different professors for other substantive first-year courses, or the impact of different kinds of first-year exams, affected student performance and thus affected the results of these studies. Perhaps the results are simply due to the fact that the intervention students had to do something more than simply orally respond to questions and the same results would be achieved

---

162. *See supra* Part V.D (reporting that the interventions did not impact students' grades in other law classes that spring).

163. *See* Henderson, *supra* note 19 (finding that the LSAT and in-class timed exams both have a speededness component).

164. *See supra* text accompanying notes 34-35.

165. *See* Henderson, *supra* note 19 (suggesting that take home essay exams without the speededness component may decrease the correlation between first-year exam grades and LSAT scores).

from any form of non-oral assignments. Likewise, any one of the feedback methods may have been what made the difference. Future studies could and should try to isolate these variables.

Of course, there may be variables we have not even considered that affected the outcome. Given that law school professors' main concern is teaching their substantive course, it will likely be impossible to design a completely unassailable scientific comparison to measure the effectiveness of a particular teaching innovation. Nonetheless, this study indicates that we can develop a cost-effective way[166] to measure the impact of various teaching methods in a classroom setting[167] and we think it suggests that law professors should consider the role of metacognition when teaching analytical skills.

VII. CONCLUSION

The Best Practices and Carnegie Foundation studies challenge law professors to examine what and how we teach and how our teaching methodologies fit with our goals of producing well-educated and capable new lawyers.[168] These theoretical works lay the groundwork for thinking broadly about what we are doing and why we are doing it. This study moves from the theoretical to the practical. In this study, we have sought to begin to fill a void in the literature regarding empirical evidence relating to specific teaching methodologies and relating to the impact of practice writing exercises on essay exam performance. We found that we could develop a cost-effective model to measure whether practice writing exercises affected essay

---

166.   The intervention professor gave the writing exercises as part of her regular course teaching. There was extra time involved during the semester in drafting the problems and model answers. There was also a little more administrative time involved in collecting and organizing the various papers. Grading and commenting on one paper during the semester took a significant amount of time (about fifteen hours total). Additionally, the study's professors also spent time collaborating during the semester to decide on substantive areas to test, and then collaborating on the use of materials and methods of teaching those substantive areas. At the end of the semester, the professors spent time developing a joint exam and rubric (although this only took slightly more time than developing an individual exam question and grading rubric). The most significant time commitment involved grading twice as many essay exam questions and collaborating during the process to ensure a uniform application of the rubric. Thus, the intervention professor spent about fifteen to twenty extra hours developing the exercises and feedback and grading one of the papers. The additional grading work consumed about another twenty hours for each professor. Admittedly, this is a significant time commitment.

167.   Others duplicating this study could follow the collaborative model we used. Or, one professor could teach two sections of the same substantive course, giving the interventions only in one section. Alternatively, a professor wishing to give the interventions and measure their effect could simply ask a colleague to collaborate on the same exam questions and the presentation of the substantive material that would be the subject of the exam questions. The collaborating professor need not get involved in either grading the exams or analyzing the results. Instead, the intervention professor could make copies of her colleagues' students' exam answers and blind grade all exams from both classes herself.

168.   STUCKEY ET AL., *supra* note 1; SULLIVAN ET AL., *supra* note 1.

exam performance. Scholars may use this same methodology in further explorations of the value of writing exercises, or in empirically measuring the effectiveness of some other teaching methodology.

This study's results indicate that student learning outcomes with regard to a particular skill subset can be empirically tested. This study provides support for, although it does not definitively prove, that other professors were correct in their assumptions that students learn better when given opportunities to practice a skill and receive feedback on that practice, and that this extends to students' ability to clearly organize and write answers to essay questions. However, the study calls into question whether practice exercises combined with feedback hold the same benefit for all students. It suggests that those who want to improve all students' performance should consider combining metacognitive exercises with whatever innovative methodology they seek to use. This area is also ripe for further study simply in terms of our traditional teaching methodologies. For example, would it improve student's performance if professors gave students metacognitive exercises designed to help them understand how and why the in-class dialogue relates to the big picture analytical skills development professors expect students to learn? This study provides a model for testing that proposition.

We realize that this study is only a very small first step in the attempt to determine the efficacy of different law school teaching methodologies empirically. However, the study indicates that we can empirically measure the effectiveness of particular teaching methodologies. Of course, especially initially, the empirical results are not beyond scientific reproach, nor even conclusive. However, as is the case with all social science research, the first step is developing a model that can be replicated and improved upon. That is what we hope to have accomplished.