2004

The LSAT, Law School Exams and Meritocracy: The Surprising and Undertheorized Role of Test-Taking Speed

William D. Henderson
Indiana University Maurer School of Law, wihender@indiana.edu

The LSAT, Law School Exams, and Meritocracy: The Surprising and Undertheorized Role of Test-Taking Speed

William D. Henderson*

Within the field of psychometrics, it is widely acknowledged that test-taking speed and reasoning ability are separate abilities with little or no correlation to each other. The LSAT is a univariate test designed to measure reasoning ability. Test-taking speed is assumed to be an ancillary variable with a negligible effect on candidate scores. This Article explores the possibility that test-taking speed is a variable common to both the LSAT and actual law school exams. This commonality is important because it may serve to increase the predictive validity of the LSAT. The author obtained data from a national and a regional law school and followed the methodology of a typical LSAT validity study, with one important exception: student performance was disaggregated into three distinct testing methods with varying degrees of time pressure: (1) in-class exams, (2) take-home exams, and (3) papers. Consistent with the hypothesis, the data showed that the LSAT was a relatively robust predictor of in-class exams and a relatively weak predictor of take-home exams and papers. In contrast, undergraduate GPA (UGPA) was a relatively stable predictor of all three testing methods.

The major implication of this study is that the current emphasis on time-pressured law school exams increases the relative importance of the LSAT as an admission criterion. Further, because the performance gap between white and minority students tends to be larger on the LSAT than UGPA (the other important numerical admissions criteria), heavy reliance on time-pressured law school exams is likely to have the indirect effect of making it more difficult for

---

* Associate Professor of Law, Indiana University School of Law—Bloomington. I would like to thank the Law School Admission Council (LSAC), which provided funding for this project. However, the opinions and conclusions contained in this Article are those of the author and do not necessarily refleet the position or policy of the LSAC. Deborah L. Schnipke of Virtual Psychometrics, LLC, provided invaluable assistance in the preparation and analysis of the data set. Michael Heise, William Kidder, James Lindgren, James Gillespie, Emile Karafiol, Ken Dau-Schmidt, David Snyder, Laura Beth Nielsen, Jeffrey Stake, Matt Hinerfeld, and Eugene Kontrovich provided valuable comments at various stages of this project. The analysis also benefited from workshops at the American Bar Foundation (ABF) and Chicago-Kent College of Law. I would like to thank Harold Krent and Katherine Baker, who made it possible for me to work full-time on this project during my visiting appointment at Chicago-Kent. Finally, I would also like to thank the deans, registrars, and various support staff at the two law schools that participated in this study. Although they remain anonymous, *see infra* note 113, they were an indispensable part of making this project possible.

minority students to be admitted through the regular admissions process. The findings of this study also suggest that when speed is used as a variable on law school exams, the type of testing method, independent of knowledge and preparation, can change the ordering (i.e., relative grades) of individual test-takers. The current emphasis on time-pressured law school exams, therefore, may skew measures of merit in ways that have little theoretical connection to the actual practice of law. Finally, this study found some preliminary evidence that the performance gap between white and minority students may be smaller on less time-pressured testing methods, including blind-graded, take-home exams. Definitive evidence on this issue will require a larger sample size.

I. Introduction

The Law School Admission Test (LSAT) is a cultural lightning rod. While some prominent scholars attack the test as a poor predictor of law school success that is biased in favor of the privileged,[1] others praise it as a valuable tool for social mobility.[2] With each admissions season, the LSAT also creates a raft of winners and losers, as acceptance letters[3] and scholarship money[4] often turn on relatively small differences in test scores. Integrally related to this process is the ranking of law schools by *U.S. News & World Report*.[5] Despite a methodology that attempts to consider a variety of substantive factors, including faculty reputation, library resources, faculty-student ratios, and bar passage, these rankings move in virtual lockstep with a school's median LSAT score.[6] Because students, legal employers, and

---

1. *See, e.g.*, Richard Delgado, *Official Elitism or Institutional Self-Interest? 10 Reasons Why the UC-Davis Should Abandon the LSAT (and Why Other Good Law Schools Should Follow Suit)*, 34 U.C. DAVIS L. REV. 593, 600–02 (2001) (arguing for the abolition of the LSAT because the LSAT explains only 16 percent of the variation in first-year grades and noting the high correlation between family income and LSAT scores); Lani Guinier, *Race Shows the Way*, LEGAL TIMES, Sept. 16, 2002, at 58, 59 (noting that the LSAT is only "9 percent better than random in predicting first-year law school grades" and that "excessively weighting the LSAT" is the true crux of the Michigan affirmative action case). *But see infra* note 53 (referring to validity studies showing the LSAT to be the best single predictor of law school grades).

2. *Wrong Way*, DAILY TEXAN, Sept. 28, 1998, at 4 (citing the dean of The University of Texas School of Law for the position that the LSAT gives students who were forced to balance a job and studies during college a "second chance" to demonstrate their potential).

3. *See* William C. Kidder, *The Rise of Testocracy: An Essay on the LSAT, Conventional Wisdom, and the Dismantling of Diversity*, 9 TEX. J. WOMEN & L. 167, 193 (2000) (comparing the dramatic dropoffs in admissions rates at four University of California law schools based on a 95 point drop in LSAT scores).

4. *See, e.g.*, Jay Conison, *Financial Management of the Law School: Costs, Resources, and Competition*, 34 U. TOL. L. REV. 37, 37–38 (2002) (stating, as dean of Valparaiso University School of Law, that "many more schools are investing far more money—through scholarships and other means—to recruit students with high LSATs"); Howard O. Hunter, *Thoughts on Being a Dean*, 31 U. TOL. L. REV. 641, 643 (2000) (noting, as dean of Emory University School of Law, that LSATs "tie in closely with scholarships").

5. *Best Graduate Schools: Schools of Law*, U.S. NEWS & WORLD REP, Apr. 15, 2002, at 64–68.

6. STEPHEN P. KLEIN, PH.D. & LAURA HAMILTON, PH.D., THE VALIDITY OF THE U.S. NEWS AND WORLD REPORT RANKINGS OF ABA LAW SCHOOLS 12 *available at* http://www.aals.org/validity.html (last visited Jan. 21, 2004) (arguing that "90% of the overall differences in ranks among schools can be explained solely by the median LSAT score of their entering classes").

alumni are often swayed by these rankings, competition between law schools has become an LSAT "arms race."[7] Although many within the legal academy lament the "overreliance on the LSAT,"[8] law faculties have generally been unwilling to bear the consequences of taking a different path, at least by themselves.[9] As one law school dean aptly noted, the situation has become a "classic 'prisoners dilemma.'"[10]

The LSAT also presents a special set of problems for minority students, who have historically posted significantly lower scores than their white counterparts.[11] If the Supreme Court's recent decision in *Grutter v. Bollinger*[12] had struck down the use of racial preferences in law school admissions, it is at least plausible that the legal academy would have finally mustered the collective will to confront its own admissions practices.[13] However, by upholding the use of racial criteria, *Grutter* ensures that a "critical mass"[14] of minority students will continue to enroll in the nation's most prestigious law schools, thus partially offsetting the effects of the LSAT test score gap. Yet, according to the Court's own language, this method of achieving racial diversity may be unlawful in 25 years.[15] Thus, insofar as the

---

7. Abiel Wong, Note, *"Boalt-ing" Opportunity?: Deconstructing Elite Norms in Law School Admissions*, 6 GEO. J. POVERTY L. & POL'Y 199, 239–41 (1999).

8. Deirdre Shesgreen, *Schools Look at the 'Whole Person,'* LEGAL TIMES, Jan. 13, 1997, at 2, 17 (quoting a Virginia law professor during an AALS panel discussion and reporting a "consensus" among conference participants that the *U.S. News & World Report* rankings had caused law schools to "put too much weight on the LSAT in their admissions decisions").

9. *See, e.g.*, ANDREA GUERRERO, SILENCE AT BOALT HALL: THE DISMANTLING OF AFFIRMATIVE ACTION 137–38 (2002) (chronicling student proposals at Boalt Hall to de-emphasize the LSAT in order to achieve diversity and faculty opposition because of the effects it might have on the law school's ranking; noting comments among some faculty that their willingness might change if students "get Harvard to change its policy first").

10. Terry Carter, *Rankled by the Rankings*, ABA J., Mar. 1998, at 46, 49 (quoting the dean of Boston University School of Law).

11. LAW SCHOOL ADMISSION COUNCIL, MINORITY DATABOOK 14 tbl.IV-1 (2002) [hereinafter LSAC MINORITY DATABOOK] (reporting the annual mean LSAT scores by ethnic subgroup for seven consecutive years with the following ranges: African American, 141 to 142; Native American, 147 to 148; Asian American, 150 to 151; Caucasian, 151 to 152; Hispanic, 146 to 147; Mexican American, 146 to 147; Puerto Rican, 137 to 138).

12. 123 S. Ct. 2325 (2003) (holding that the educational benefits of a diverse student body at a public law school is a compelling state interest and upholding the admissions policy of the University of Michigan Law School).

13. *See* Shesgreen, *supra* note 8, at 17 (quoting the dean of Northeastern University School of Law, "Is this crisis [in racial preference policies] an opportunity to rethink the whole admissions paradigm?").

14. *Grutter*, 123 S.Ct. at 2339 (noting that "the Law School's concept of critical mass is defined by referenee to the edueational benefits that diversity is designed to produce").

15. *Id.* at 2347 ("We expect that 25 years from now, the use of racial preferences will no longer be necessary to further the interest approved today."); *id.* at 2350 (Thomas, J., concurring in part and dissenting in part) ("I agree with the Court's holding that racial discrimination in higher edueation admissions will be illegal in 25 years."). *But see* CIVIL RIGHTS PROJECT, HARVARD UNIV., REAFFIRMING DIVERSITY: A LEGAL ANALYSIS OF THE UNIVERSITY OF MICHIGAN AFFIRMATIVE ACTION CASES (2003), *available at* http://www.civilrightsproject.harvard.edu/

Court has sheltered law schools from the cultural and political fallout of their LSAT-driven admission policies, *Grutter* may be less of a solution than a reprieve.

This Article opens up a new avenue of research with the potential to break the LSAT gridlock. Based on a detailed sample set drawn from a national and a regional school, this Article presents strong empirical evidence that test-taking speed is a variable that affects student performance on both the LSAT and actual law school exams. The issue of speed is important because the LSAT is designed to provide a standard measure of a candidate's reasoning skills.[16] Although it may be intuitively appealing to assume that speed is simply a component of a candidate's underlying reasoning ability (i.e., I.Q., fluid intelligence, "g"[17]), this view lacks empirical support. Within the field of psychometrics, test-taking speed and reasoning ability are viewed as distinct, separate abilities with little or no correlation.[18]

Because speed and reasoning skills are viewed as distinct and separate abilities, testing theorists often distinguish between tests that measure "power" and tests that measure "speed."[19] Like most tests that measure

---

> This sentence [regarding the 25 year provision] should be construed as the Court's dictum expressing, by reference to the passage of time since the Bakke deeision, its aspiration—and not its mandate—that there will be enough progress in equal education opportunity that raee-conscious policies will, at some point in the future, be unnecessary to ensure diversity.

Id.

16. *See* LAW SCHOOL ADMISSION COUNCIL, LAW SCHOOL ADMISSION REFERENCE MANUAL 2003–2004, at 7 (2003) [hereinafter LSAC REFERENCE MANUAL] ("The LSAT provides a standard measure of acquired reading and verbal reasoning skills."); LAW SCHOOL ADMISSION COUNCIL, THINKING ABOUT LAW SCHOOL: A MINORITY GUIDE 10–11 (1995) ("The LSAT is designed to measure your reasoning skills, so there are no shortcuts or ways to 'beat' the test.").

17. *See generally* PAUL KLINE, A PSYCHOMETRIC PRIMER ch. 5 (2000) (providing an overview of intclligence and ability tests, including terminology such as I.Q., fluid intelligence, and "g"). Kline comments that the "correlation between intelligence tests and academic attainment can be explained most parsimoniously by conceptualizing intelligence as the ability to reason." *Id.* at 79.

18. *See, e.g.*, *id.* at 150–51 (2000) (noting that "all rccent research shows [mental speed] to be only modestly correlated with intelligence" and that "[i]t can be concluded that speed and intelligence are separate"); Lazar Stankov et al., *Models and Paradigms in Personality and Intelligence Research*, *in* INTERNATIONAL HANDBOOK OF PERSONALITY AND INTELLIGENCE 25–28 (Donald H. Saklofske & Moshe Zeidner, eds., 1995) (noting that "speed in doing difficult (power) tests shows zero correlation [with intelligence]"); JOHN B. CARROLL, HUMAN COGNITIVE ABILITIES: A SURVEY OF FACTOR-ANALYTIC STUDIES 506–08 (1993) ("Empirically, pure (or nearly pure) measures of speed ['time or rate of performance'] and level abilities ['level of task difficulty at whieh an individual can perform with a specified amount of accuracy'] tend to have very low or even zero intercorrelations."); P.O. White, *Individual Differences in Speed, Accuracy and Persistence: A Mathematical Model for Problem Solving*, *in* THE MEASUREMENT OF INTELLIGENCE 246, 253–58 (H.J. Eysenck ed., 1973) (finding in a study of 93 test-takers on Raven's progressive matrices test (a widely used I.Q. test), that speed was "uncorrelated with accuracy (r = -0.01) and with I.Q. (r = -0.06), but . . . negatively correlated with persistence (r = -0.35)").

19. *See, e.g.*, CARROLL, *supra* note 18, at 440–46 (discussing differences between "speed" and "power" as commonly used by psychometricians).

reasoning ability, the LSAT is essentially designed to be a power test.[20] However, recent research has suggested that the LSAT may be much more "speeded" than originally believed.[21] Similarly, within the legal academy, it is widely acknowledged that the typical law school exam is given under significant time pressure.[22] Although the time component of the LSAT is subject to various psychometric parameters, the time pressure on law school exams is not grounded in any formal or well defined theory.[23] Rather, it is driven by a variety of nontheoretical factors, including the need to generate a grading curve, the desire to limit the volume of words that must be read, and the simple adherence to tradition.[24]

This common element of time pressure is noteworthy because LSAT validity studies are based on the test's ability to predict law school performance, particularly during the first year.[25] This study is designed to

---

20. *See* DEBORAH L. SCHNIPKE & DAVID J. SCRAMS, MODELING ITEM RESPONSE TIMES WITH A TWO-STATE MIXTURE MODEL: A NEW APPROACH TO MEASURING SPEEDEDNESS 2 (LSAC, Computerized Testing Report 96-02, 1999) [hereinafter TWO-STATE MIXTURE MODEL] ("Most aptitude and assessment tests are designed essentially as power tests .... The Law School Admission Test (LSAT) is such a test.").

21. *See infra* notes 65–79 and accompanying text.

22. *See, e.g.*, Philip C. Kissam, *Law School Examinations*, 42 VAND. L. REV. 433, 438 (1989) (noting that "the concept of a time constraint is the most widely shared characteristic of law school exams"); *see also infra* note 85 and accompanying text.

23. In preparation for this Article, the only explicit justification for time-pressured law school exams I found was the following passage from a law school exam study aid written by a University of Southern California law professor:

> Students often criticize law school examinations for their immense time pressure. Frankly, attorneys often work under severe time constraints, as any student who clerks in a law office can attest. For that reason I think it is fundamentally fair to expect law students to respond quickly and cogently on examinations. But I am a law professor, and so naturally I defend the time pressure on law school exams.

CHARLES H. WHITEBREAD, THE EIGHT SECRETS OF TOP EXAM PERFORMANCE IN LAW SCHOOL 9 (1995).

Although Whitebread is probably right that lawyers often work under severe time constraints, it is also true that good lawyers need to communicate clearly and with precision. Anyone who has ever graded a three-hour essay exam knows that it is a singularly unpleasant experience, primarily because of poor organization and prose. My colleague has described it as "counting spotted fish in a barrel." This outcome is due to time pressure. *See infra* note 126 (reporting the observation of a registrar at a national law school that the professors who gave take-home exams preferred them because they produced better written and organized answers); Ruthann Robson, *The Zen of Grading*, 36 AKRON L. REV. 303, 311 (2003) (discussing the dread of grading Blue Book exams and noting that "individual research papers or ... take-home exams are such a joy to read by comparison"); SCOTT TUROW, ONE L 168 (Warner Books 1997) (1977) (discussing the "importance of time" on law school exams and noting that four hours was "not a quarter of the time I'd need to frame a reasonably thorough response"); *see also infra* subpart IV(C).

24. *See* Kissam, *supra* note 22, at 438, 453 (asserting, from his viewpoint as a law professor at the University of Kansas, that time-constrained examinations are administered for "no particular reason, other than tradition"). For a discussion of how "speed" might matter in the legal profession, see *infra* subpart V(B).

25. *See* LSAC REFERENCE MANUAL, *supra* note 16, at 11 (noting that "most law schools have participated in studies that have compared students' LSAT scores with their first-year grades in law

ascertain whether time pressure (or, stated in the alternative, test-taking speed) is a variable common to both the LSAT (the predictor variable) and student grades (the criterion variable). If this relationship is present, then part of the predictive validity of the LSAT is likely attributable to test-taking speed rather than reasoning ability.[26] Using student performance data supplied by an elite national law school and a middle-level regional law school, this study utilizes a methodology similar to LSAT validity studies, with one important exception. The study disaggregates student performance on three distinct testing methods with varying degrees of time pressure: (1) in-class exams; (2) take-home exams; and (3) paper assignments.[27]

Consistent with the hypothesis on test-taking speed, the LSAT was a relatively robust predictor of performance on in-class exams but a relatively weak predictor of performance on take-home exams and paper assignments.[28] In contrast, the predictive power of undergraduate grade point average (UGPA) remained much more stable on all three testing methods.[29] The findings of this study strongly suggest that, within each sample population, a significant portion of the predictive power of the LSAT appears to be based on differences in test-taking speed rather than reasoning ability.[30]

The upshot of the data is *not* that the LSAT is a flawed predictor of law school success.[31] Rather, this Article provides strong empirical evidence that the predictive validity of the LSAT is inextricably tied to the legal academy's heavy and under-theorized reliance on time-pressured exams. If the results from the national and regional law schools in this study are fairly representative of other law schools, a migration away from time-pressured, in-class exams and toward take-home exams and papers would (a) reduce the

---

school" and that "these studies show that LSAT scores help to predict which students will do well in law school"); *see also* sources cited *infra* note 53.

26. The mechanics of this relationship are explained in Part II and subpart V(A), *infra* (both discussing the effect of test-taking speed on both the LSAT and law school exams).

27. The project methodology is set forth in detail in subpart III(B), *infra*.

28. *See infra* subpart V(A).

29. The relative stability of UGPA on all three testing methods (in-class exams, take-home exams, and papers) is significant because it rebuts the criticism that grades on take-home exams or papers are more subjective or less likely to produce a discernible grading curve. *See infra* note 123.

30. *See infra* subpart V(A).

31. The LSAT is designed to predict law school performance, a task it performs better than any other single predictor. *See* sources cited *infra* note 51. The significance of this study is that it demonstrates that testing method (in-class exams, take-home exams, or papers) is an independent variable that affects the ordinal ranking (i.e., grades) of law students. Without empirical evidence, other academics have flatly stated that testing method is irrelevant. *See infra* note 180. As the data shows, the LSAT is a much more robust predictor of performance during in-class exams than during take-home exams and papers, which are obviously less time pressured. However, the large proportion of in-class exams given at most law schools may be what underlies the high correlation between LSAT scores and law school grades that emerges in all LSAT validity studies. *See infra* subpart IV(B). In other words, the LSAT predicts best on the most common testing method. This is hardly a flaw. If the LSAT were *less* speeded, it would likely be a *worse* predictor of in-class exams. *See infra* note 81 and accompanying text. This same effect cuts the other way: a less speeded LSAT might be a better predictor of performance on papers and take-home exams.

relevance of the LSAT as an admission criterion and (b) increase the relative importance accorded to UGPA. A reweighting of admission criteria in favor of UGPA will have the effect of admitting more minority candidates through the regular admissions process.[32]

The lack of a clearly defined theoretical basis for time-pressured law school exams is also problematic because of the distributional effects it has on students. When test-taking speed is a variable on law school exams, the type of testing method, independent of knowledge or preparation, can change the ordering (i.e., relative grade) of individual test-takers. Students with fast test-taking speed are clearly favored on in-class exams.[33] Since numerous academic and career opportunities often hinge on relatively small variations in law school grades, test-taking speed implicitly becomes a measure of merit. Yet, it could certainly be argued that papers and take-home exams are a much closer analogue to the practice of law, in terms of both time pressure and the creation of a final work product that might be relied upon by a client, another lawyer, or a court.[34]

Finally, aside from excessive weighting of the LSAT, the emphasis on time-pressured law school exams may have a disproportionately large impact on minority groups. Several earlier studies have found evidence that minority students may be disparately affected by stringent time limits on exams.[35] Despite the constraint of a relatively small sample size, this study

---

32. *See* GUERRERO, *supra* note 9, at 170–71 (presenting a table showing that all minority groups would fare significantly better under a "UGPA only" admissions model than a combined "LSAT/UGPA numbers only" model, while whites would fare marginally worse); Danny Holley & Thomas Kleven, *Minorities and the Legal Profession: Current Platitudes, Current Barriers*, 12 T. MARSHALL L. REV. 299, 309–10 (1987) (reviewing empirical evidence of minority admissions and finding that the LSAT "ha[s] a far greater negative impact than UGPA"); Linda F. Wightman, *The Threat to Diversity in Legal Education: An Empirical Analysis of the Consequences of Abandoning Race as a Factor in Law School Admission Decisions*, 72 N.Y.U. L. REV. 1, 8 (1997) (commenting, based on her experience as a psychometrician for the LSAC, that "the discrepancy between applicants of color and white applicants is larger for LSAT scores than for UGPAs"). For a discussion of what I mean when I use the term "regular admission process," see *infra* note 171.

33. Subpart II(A) provides a stylized example of how testing method can skew student grade averages.

34. *Cf.* Judge Alex Kozinski, *Foreword to* EUGENE VOLOKH, ACADEMIC LEGAL WRITING: LAW REVIEW ARTICLES, STUDENT NOTES, AND SEMINAR PAPERS (2003) ("It is difficult to overstate the importance of a written paper for a young lawyer's career . . . . Writing a paper engages so much of a lawyer's art that no other predictor of likely success on the job comes close."). For a discussion of "speed" and the practice of law, see *infra* subpart V(B).

35. *See, e.g.*, DEBORAH L. SCHNIPKE & PETER J. PASHLEY, ASSESSING SUBGROUP DIFFERENCES IN ITEM RESPONSE TIMES 2 (LSAC, Computerized Testing Rep. 97-03, 1999) (citing several studies on the tendency of African American and Hispanic test takers to spend more time on each test item and thus, the likelihood that they will reach slightly fewer items than white test takers; commenting that "[t]hese studies suggest that time limits on tests may differentially affect test scores of some subgroups"); LINDA F. WIGHTMAN & DAVID G. MULLER, COMPARISON OF LSAT PERFORMANCE AMONG SELECTED SUBGROUPS 6 (LSAC, Statistical Rep. 90-01, 1990) (examining the tendency of ethnic subgroups to answer every question on the LSAT and noting that "[e]vidence of speededness is particularly dramatic for members of the black, Puerto Rican, and Hispanic subgroups"); Franklin R. Evans & Richard R. Reilly, *A Study of Speededness as a Source*

also found limited preliminary evidence that the performance gap between white and minority students may be smaller on take-home exams and papers than on in-class exams.[36] The lack of a well-defined nexus between test-taking speed and the practice of law is therefore especially troublesome. If additional evidence suggests that time-compressed, in-class exams have a disparate impact on one or more ethnic subgroups, the legal academy will need to put forward a fairly rigorous theoretical basis to justify a testing regime in which a large proportion of first-year grades (regional law school, 80.6%; national law school, 74.9%)[37] and cumulative law school grades (regional law school, 74.6%; national law school, 61.3%)[38] rely on this testing method.

This Article is primarily a hypothesis test on whether test-taking speed is a variable that is common to both the LSAT and actual law school exams. Its main thrust is empirical rather than normative. Therefore, the structure of this Article largely follows a social science format. Part II places this study into context by framing the research question and reviewing relevant background literature. Part III describes the sample group and project methodology. Part IV presents the project results. Finally, Part V discusses the implications of this study for law school testing and outlines a preliminary theoretical framework for predicting and interpreting law school performance.

II. Background on Speed and Reasoning Ability

Discussions on standardized tests inevitably have to grapple with some unflattering history. For example, critics of the LSAT frequently point out that the test descends from a long line of intelligence tests, which were originally developed by eugenicists who believed in the intellectual superiority of Caucasian men.[39] Because these tests continue to reinforce negative racial stereotypes, some critics argue that they are culturally biased

---

of Test Bias, 9 J. EDUC. MEASUREMENT 123, 127 (1972) (reporting evidence of proportionately larger increases in the LSAT seores for black versus white candidates when the reading comprehension portion of the LSAT was given under conditions in which the speed of test taking was not a factor).

36. *See infra* section IV(C)(4). It is noteworthy that take-home exams in both the regional law school and national law school sample groups were blind graded; this rebuts suggestions of racial favoritism.

37. *See infra* tbl.17 (describing national law school testing methods); tbl.20 (describing regional law school testing methods).

38. *See id.*

39. *See, e.g.*, Delgado, *supra* note 1, at 595–96 (discussing the genealogy of the current system of standardized testing in higher education and claiming that Carl Campbell Brigham, an early president of the College Board, was "an unapologetic race-purifier for whom Catholics, Jews, and east Europeans represented a defective strain of humanity"); GUERRERO, *supra* note 9, at 12 ("The LSAT [was] derived from IQ (intelligence) tests designed by eugenicists in the early part of the century, who believed that the test measured a biologically grounded, genetically inherited quality that was tied to ethnicity." (citing NICHOLAS LEMANN, THE BIG TEST 33 (1999))).

and should be abandoned.[40] Unfortunately, these arguments tend to overlook the significant changes that the LSAT has undergone during its nearly six decade existence,[41] including substantial and ongoing efforts to identify and eliminate sources of gender and ethnic bias.[42] Ultimately, what animates many criticisms of the LSAT is an underlying dissatisfaction with the test's distributional consequences as it exists today.

In an effort to breakup a logjam on a contentious social issue, this Article takes a more pragmatic approach. Specifically, the LSAT, as well as other standardized tests, has become firmly embedded within the fabric of American higher education. Because this system creates both winners and losers, arguments that emphasize bias or perceived unfairness are unlikely to alter the status quo. Rather than criticizing the *outcomes* that are produced by the LSAT, this Article focuses on the LSAT itself, on its own terms, including its relationship to actual law school testing methods. The advantage of this approach is that it can demonstrate the technical limitations of the LSAT while simultaneously identifying a broader range of options to help us balance the dual goals of meritocracy and social justice. At the end of the day, the academy is given a brand new tool to deal with an immensely difficult problem.

Part I of this Article provides relevant background information that frames the important distinction between test-taking speed and reasoning ability. Subpart A provides two hypothetical examples that illustrate how the variable of test-taking speed could affect performance on (1) the LSAT, and (2) law school exams. Subpart B then reviews the psychometric literature on the relationship between test-taking speed and reasoning ability.

A. Research Question

This study examines the question of whether test-taking speed[43] is a variable that affects performance on both the Law School Admission Test

---

40. *See, e.g.*, Delgado, *supra* note 1, at 594–97.

41. *See generally* William P. LaPianna, *A History of the Law School Admission Council and the LSAT*, Keynote Address, 1998 LSAC Annual Meeting; Thomas O. White, *LSAC/LSAS: A Brief History*, 24 J. LEG. EDUC. 369 (1984).

42. *See* LSAC REFERENCE MANUAL, *supra* note 16, at 59 (noting that the LSAC "applies two procedures to ensure that the LSAT is fair to all test takers regardless of racial, ethnic, gender, regional, or national background," including an item-by-item review by trained reviewers in which questions unfair or offensive to subgroups are eliminated, and identification and elimination of questions which produce large statistical disparities among subgroups that are inconsistent with the test as a whole).

43. Note that the construct here is "test-taking speed." In the psychometric literature, test-taking speed refers to rate of work in responding to relatively complex test questions, which is the relevant measure here; it is conceptually and empirically distinct from measures of cognitive speed that focus on response times (RT) or inspection times (IT) of elementary cognitive tasks (ECT), which have been shown to be modestly correlated with intelligence. For discussions that distinguish different strands of research on cognitive speed see CARROLL, *supra* note 18, at 475–78, 506–09 and Stankov et al., *supra* note 18, at 25–28 (noting that "speed in doing difficult (power) tests shows zero correlation [with 'intelligence']").

(LSAT) and actual law school exams. The core psychometric fact that underlies this project is that test-taking speed and reasoning ability, which is the construct the LSAT is designed to measure,[44] are distinct and separable abilities that have little or no correlation to each other.[45] As a result, testing theorists often distinguish between tests that measure "power" and tests that measure "speed."[46] In order to avoid a merging of these two constructs, test makers generally pay close attention to time limitations. If a large proportion of test takers are unable to reach and attempt each question, the results of a power test may be confounded by a "speededness" variable.[47]

The LSAT is designed to measure "acquired reading and verbal reasoning skills."[48] This construct is presumably relevant to law school performance, particularly during the first year. Although there is some empirical evidence[49] and anecdotal evidence from LSAT study aids[50] that the LSAT may include a significant speededness component, this issue has garnered relatively little interest from researchers.

---

44. *See* sources cited *supra* note 16.

45. *See supra* note 18.

46. For discussions distinguishing between tests that measure power and tests that measure speed see, CARROLL, *supra* note 18, at 440–46, and TWO-STATE MIXTURE MODEL, *supra* note 20, at 2.

47. *See infra* note 64 and accompanying text.

48. LSAC REFERENCE MANUAL, *supra* note 16, at 7.

49. TWO-STATE MIXTURE MODEL, *supra* note 20, at 1–2, 14–16 (noting that the LSAC current measure of speededness "probably underestimates the true amount of speededness on the [LSAT]," and analyzing data from a nonadaptive computerized administration of the GRE, which closely parallels the logical reasoning and analytical reasoning of the LSAT, and finding a high degree of speededness); Evans & Reilly, *supra* note 35, at 117 (reporting a study from the early 1970s finding that LSAT test takers consistently "score higher under unspeeded conditions"); Matthew Ben Hinerfeld, The LSAT: Test Speededness and Other Failings of the Law School Admission Council/Law School Admission Service 33–34 & nn.89–90 (Supervised Analytic Writing, Yale Law School, Feb. 3, 1992) (unpublished manuscript on file with author) (noting that at least one member of the LSAC Test Development Committee acknowledged that the error rates tend to go up at the end of each LSAT section, while other members noted that many test takers are unable to finish sections of the test). Committee member David Kaye of Arizona State University School of Law has stated:

> There is concern about the speededness of the test, because there is evidence that a fair percentage of people don't finish each section. The rate at which they get [the answers] right drops off as they go along . . . . [The 1991 change] was expected to ameliorate the problem. I doubt it's going to cure it.

Hinerfeld, *supra*, at 34 (alterations in original).

50. *See, e.g.*, THOMAS H. MARTINSON, LSAT 2002: TEACHER-TESTED STRATEGIES AND TECHNIQUES FOR SCORING HIGH 12 (2002) ("The scoring mechanism for the LSAT is the simple formula 'score = correct answers' . . . . There is a trade-off between speed and accuracy, one that only you can find through practice."); ADAM ROBINSON & ROB TALLIA, PRINCETON REVIEW CRACKING THE LSAT 3 (2002) ("Very few people get a perfect score, mainly because they're not given enough time to answer all the questions."); THOMAS O. WHITE, LSAT SUCCESS 63 (1998) ("Of the many factors that determine the LSAT score, time is the most important one . . . . Most people find that the allotted 35 minutes is insufficient to consider each question in a test section carefully.").

One explanation for the scant attention paid to the speededness issue is the efficacy of the LSAT as it is currently administered. Numerous correlation studies spanning nearly three decades have shown that the LSAT is usually a better predictor of law school performance than undergraduate grade point average (UGPA).[51] As noted in a recent summary of LSAT correlations studies, the "LSAT alone continues to be a better predictor of law school performance than is [undergraduate GPA] alone."[52] Moreover, in the two-variable regression model used by most U.S. law schools to predict first-year performance—and thus make admissions decisions—the LSAT is usually the dominant variable.[53] Thus, even if the LSAT is affected by a speededness confound, it could be argued that correcting it will only make the LSAT a *more* accurate predictor of law school performance. Since the LSAT is already the best predictor available, researchers might reasonably assume that any improvement in predictive validity would have little practical consequence.

This study considers a possibility that goes beyond a relatively straightforward speededness confound. Specifically, this study examines the hypothesis that *both* the LSAT (the predictor variable) *and* law school exams (the criterion variable) are affected by a significant speededness variable. This study posits that performance on both the LSAT and actual law school exams are affected by at least two variables: (1) verbal reasoning and reading skills, and (2) test-taking speed. Therefore, if time constraints are relaxed on the predictor variable (LSAT) or the criterion variable (law school exams), the hypothesis predicts that the correlation between LSAT and law school grades will decline at statistically significant levels. In other words, part of

---

51. *See, e.g.*, LISA C. ANTHONY ET AL., PREDICTIVE VALIDITY OF THE LSAT: NATIONAL SUMMARY OF THE 1995-1996 CORRELATION STUDIES 14 (LSAC, LSAT Technical Rep. No. 97-01, 1999) (reviewing data from 165 law schools and noting that the "LSAT alone continues to be a better predictor of law school performance than is UGPA alone"); LINDA F. WIGHTMAN, BEYOND FYA: ANALYSIS OF THE UTILITY OF LSAT SCORES AND UGPA FOR PREDICTING ACADEMIC SUCCESS IN LAW SCHOOL 15 (LSAC, Research Rep. No. 99-05, 2000) [hereinafter WIGHTMAN, BEYOND FYA] (collecting data from 142 law schools and finding that the "LSAT tended to be a better predictor alone than UGPA alone"); LINDA F. WIGHTMAN, PREDICTIVE VALIDITY OF THE LSAT: A NATIONAL SUMMARY OF THE 1990-92 CORRELATION STUDIES 9 (LSAC, Research Rep. 93-05, 1993) [hereinafter WIGHTMAN, PREDICTIVE VALIDITY] (reviewing data from 167 law schools and reporting that "for each of the study years, the LSAT score is a substantially better predictor of first-year performance in law school than is the undergraduate grade point average"); Franklin R. Evans, *Recent Trends in Law School Validity Studies*, *in* 4 REPORTS OF LSAC SPONSORED RESEARCH, 1978–1983, 347, 359 (1984) (reviewing data from 140 law schools and reporting that the "LSAT is currently a better predictor of that performance than are undergraduate grades" and that this trend has "been observed for several years").

52. ANTHONY ET AL., *supra* note 51, at 14.

53. *See, e.g.*, ANTHONY ET AL., *supra* note 51, at 11 tbl.4, 14 (reviewing first-year correlation data from 165 law schools over a two-year period and reporting that "the median validity coefficient for LSAT alone is .41, compared with .26 for UGPA alone"); WIGHTMAN, BEYOND FYA *supra* note 51, at 19 tbl.6 (summarizing in table form standardized regression weights (i.e., beta weights) for both first-year and cumulative law school grades from 142 law schools and showing that the LSAT had a larger weighting for all school clusters during all time periods).

the predictive validity of the LSAT may be attributable to test-taking speed rather than a loading of acquired verbal reasoning and reading skills, which is the construct the LSAT is designed to measure. If this situation is true, we are then confronted with the more difficult question of whether there is a sound educational, theoretical, or empirical basis for emphasizing speed on law school exams.[54]

To illustrate the potential relationship between (1) test-taking speed and the LSAT, and (2) test-taking speed and law school performance, I have provided two stylized examples. In the first example, which uses the "rapid-guessing behavior/solution behavior" terminology developed by Schnipke and Scrams,[55] two law school candidates, Anderson and Benton, take the same LSAT exam on the same day.

Table 1: Hypothetical Comparison of Two Students with a 165 LSAT Score

LSAT Exam	*Anderson*	*Benton*
Number Attempted	101	80
Answered Correctly	82	78
Accuracy	81.7%	97.5%
Number of Rapid Guesses	0	21
Number Right Through Rapid Guesses	0	4
Raw Score	82	82
LSAT Score	165	165
Percentile	94th percentile	94th percentile

As summarized in Table 1, Anderson reaches all 101 questions, and thus is able to engage in "solution behavior" for all test items. Anderson answers 82 questions correctly, which results in an LSAT score of 165 (~94 percentile). When engaging in solution behavior, Anderson's accuracy is 81.7%. In contrast, Benton reaches only 80 questions, answering 78 correctly. However, during the last minute of each section, Benton fills in the remaining unanswered questions ("rapid-guessing behavior"). Getting an

54. For a discussion of "speed" and the legal profession, see *infra* subpart V(B).

55. For discussions delineating these two categories of test-taking behavior, see DEBORAH L. SCHNIPKE & DAVID J. SCRAMS, EXPLORING ISSUES OF TEST TAKER BEHAVIOR: INSIGHTS GAINED FROM RESPONSE-TIME ANALYSES 11 (LSAC, Computerized Testing Rep. No. 98-09) [hereinafter TEST TAKER BEHAVIOR] and TWO-STATE MIXTURE MODEL, *supra* note 20, at 1.

additional four answers correct, Benton also achieves a raw score of 82 and an LSAT score of 165. Yet, when engaging in solution behavior, Benton's accuracy is 97.5%. Because both Anderson and Benton received identical LSAT scores, the logical inference is that they possess similar abilities. However, in this example, it is apparent that Benton is relatively slow but very accurate, while Anderson is relatively fast but not quite as accurate.

The second example illustrates how differences in test-taking speed and ability might be differentially affected by law school testing methods that vary along the dimension of time. Consider hypothetical Midland Law School, which uses proctored, in-class exams for 1/3 of its classes, take-home exams for the next 1/3, and papers for the last 1/3. Three law students, Cassidy, Donaldson, and Edmonds, all have identical LSAT scores and UGPA. Upon graduation from law school, each student also has an identical 3.67 grade point average, once again raising the presumption that they have similar abilities.

Table 2: GPAs at (hypothetical) Midland Law School
When Testing Methods Are Equally Proportioned

GPA	*Cassidy*	*Donaldson*	*Edmonds*
In-Class	4.33 (A+)	3.67 (A-)	3.00 (B)
Take-Home	3.67 (A-)	3.67 (A-)	3.67 (A-)
Paper	3.00 (B)	3.67 (A-)	4.33 (A+)
Total LGPA	3.67 (A-)	3.67 (A-)	3.67 (A-)

However, as shown in Table 2, when their law school transcripts are disaggregated, the following pattern is evident. Cassidy does the best on in-class exams (4.33) and the worst on papers (3.00). Edmonds does the best on papers (4.33) and the worst on in-class exams (3.00). Finally, Donaldson does strong but not exceptional work on all three testing methods (3.67).

Table 3: GPAs at (hypothetical) Midland Law School When the Majority of Classes Use In-Class Exams

Proportion (%) In-Class/Take-Home/Paper	*Cassidy*	*Donaldson*	*Edmonds*
33 / 33 / 33	3.67	3.67	3.67
50 / 25 / 25	3.84	3.67	3.51
62 / 18 / 20	3.95	3.67	3.39
79 / 7 / 14	4.09	3.67	3.23

In this hypothetical, the identical 3.67 GPA for Cassidy, Donaldson, and Edmonds is entirely attributable to the arbitrary assumption that each testing method is given in equal 1/3 proportions. As shown in Table 3, when the proportions are more heavily weighted toward the in-class exam, which is the dominant testing method at most U.S. law schools,[56] the grade averages between the three students begin to diverge sharply. Under a regime in which 50% of all credits are earned by taking in-class exams and the remaining credits are divided between take-home exams (25%) and papers (25%), Cassidy emerges as the better student. Although Cassidy's 3.84 average may not be markedly different from Edmond's 3.51 average, a wide variety of academic and career opportunities (such as law review membership, judicial clerkships, commencement honors, and interviews with prestigious law firms) often hinge on relatively small variations in law school grades. When reliance on in-class exams is increased further, the gap between Cassidy, Donaldson, and Edmonds widens even more. In terms of the present study, the 62/18/20 testing regime reflects the proportion of each

56. *See* Adam G. Todd, *Exam Writing as Legal Writing: Teaching and Critiquing Law School Examination Discourse*, 76 TEMP. L. REV. 69, 69 n.1 (2003) ("The typical law school exam consists of one or a series of essay questions that must be answered within a time constraint ranging from two to three hours."); Robert C. Downs & Nancy Levit, *If It Can't Be Lake Woebegone . . . A Nationwide Survey of Law School Grading and Grade Normalization Practices*, 65 UMKC L. REV. 819, 822–23 (1997) ("The typical law school examination is a single, end-of-semester or end-of-course test on which most or all of the course grade is based. Usually, these are timed two to four hour essay tests, typically hand-written, either open or closed book.") (footnote omitted); Mark A. Godsey, *Educational Inequalities, the Myth of Meritocracy, and the Silencing of Minority Voices: The Need for Diversity on America's Law Reviews*, 12 HARV. BLACKLETTER L.J. 59, 76 (1995) ("The typical law school exam is a three-to-five hour affair, consisting primarily of essay questions.").

testing method found at the national law school.[57] Similarly, the 79/7/14 testing regime reflects the proportion of each testing method found at the regional law school sample.[58]

The question raised by this thought experiment is whether Cassidy is objectively the best student. All other factors being equal, the employment market will likely treat Cassidy as a more desirable candidate than Donaldson or Edmonds.[59] However, if Midland Law School migrated away from proctored, in-class exams and toward assigned papers, Edmonds would emerge as the better student, and the market would respond accordingly. Yet, a priori, there is no clear educational or theoretical basis for preferring one candidate over another.[60] A related issue is whether differential patterns of performance are potentially suggestive of a student's ability to excel in a specific type of legal practice (such as appellate practice, trial practice, or transactional practice). A third area of consideration is whether this information would potentially alter the hiring preferences or practice assignments of legal employers. For example, would a relatively low grade average on in-class exams be more readily discounted if a candidate's grades also reflected a superior writing ability? The disaggregation process may provide insights on a student's particular strengths and weaknesses.

B. Literature Review

The relationship between speed and intelligence is a topic that has produced an extensive amount of research. During the early 1900s, when aptitude and achievement tests were first being developed, theorists generally believed that speed and accuracy (i.e., the ability to select correct answers) reflected the same underlying construct of general intelligence, thus "implying that it does not matter whether a testtaker's ability is measured on the scale of accuracy (power), a scale of speed, or some combination of the two."[61] With the widespread adoption of aptitude tests, time limits became relatively commonplace, primarily for administrative convenience.[62]

---

57. *See infra* tbl.17 (describing the testing methods employed by a sample of national law schools).

58. *See infra* tbl.20 (reporting the testing methods used by a sample of regional law schools).

59. *See* Gerald F. Hess, *Heads and Hearts: The Teaching and Learning Environment in Law School*, 52 J. LEGAL EDUC. 75, 78 (2002) (noting that "[g]rades and class rank are significant gatekeepers to the reward system during and after law school—law review membership, research or teaching-assistant positions, internships, and jobs").

60. For a discussion of "speed" and the legal profession, see *infra* subpart V(B).

61. TEST TAKER BEHAVIOR, *supra* note 55, at 8 (citing the well-known CHARLES SPEARMAN, THE ABILITIES OF MAN (1927) as an example of this view); *see also* CARROLL, *supra* note 18, at 450 (noting that during "the early days of psychological testing," speed and ability were viewed as "two sides of the same coin").

62. TEST TAKER BEHAVIOR, *supra* note 55, at 8; Edward J. Morrison, *On Test Variance and the Dimensions of the Measurement Situation*, 20 EDUC. & PSYCHOL. MEASUREMENT 231, 231 (1960) ("It is something of a surprise to discover that the popularity of time-limit tests is due more to their

Because speed and accuracy were viewed as different manifestations of the same underlying construct, it was assumed that time limits were irrelevant to test results.

Since the late 1930s, a vast amount of research has demonstrated that speed and accuracy are in fact separate variables. For example, in the exhaustive analysis of every major study on cognitive speed since the early 1900s, Professor Carroll concluded that "[e]mpirically, pure (or nearly pure) measures of speed and level abilities [i.e., ability to accurately perform difficult tasks] tend to have very low or even zero intercorrelations."[63] Within the psychometric field, there is a general consensus that the results of an aptitude or assessment test can be confounded if the "speededness" component is too large.[64]

On tests of reasoning ability, such as the LSAT, speededness is usually determined by the number of unanswered questions remaining on a test.[65] If a substantial majority of test takers complete all questions within the allotted time, then the speededness component is generally assumed to be negligible.[66] However, when a test score is based only on the number of correct answers, which is the method used for the LSAT, test takers have an incentive to rapidly guess on all unreached items before time expires, or to adopt pacing strategies that compromise accuracy in order to attempt more test items.[67] Not surprisingly, in a recent report published by the LSAC,

---

practical administrative advantages and their demonstrated usefulness than to any experimentally-supported rationale governing the imposition of time limits on performance.").

63. CARROLL, *supra* note 18, at 507. Chapter 5 of Professor Carroll's book is an exhaustive summary and analysis of every major study on the issue of cognitive speed and intelligence since the early 1900s.

64. *See, e.g.*, DAVID J. SCRAMS & DEBORAH SCHNIPKE, MAKING USE OF RESPONSE TIMES IN STANDARDIZED TESTS: ARE ACCURACY AND SPEED MEASURING THE SAME THING? 11 (LSAC, Computerized Testing Rep. 97-04, May 1999) (finding that speed and accuracy were unrelated on the logical and analytical reasoning portion of the GRE and that the "test-taker ordering can change" if the speededness component is too large); CARROLL, *supra* note 18, at 507 (stating that the level of speededness can be "an influential and disturbingly misleading confound"); William G. Mollenkopf, *Time Limits and the Behavior of Test Takers*, 20 EDUC. & PSYCHOL. MEASUREMENT 223, 229 (1960) ("The same test material given under speeded conditions may not measure the same behavior as under power conditions."); Alexander G. Wesman, *Some Effects of Speed in Test Use*, 20 EDUC. & PSYCHOL. MEASUREMENT 267, 272 (1960) ("In any test which is a combination of speed and power, the role which speed plays may be a very confusing one.").

65. TWO-STATE MIXTURE MODEL, *supra* note 20, at 2 (noting that on the LSAT, "speededness is currently measured by calculating the proportion of test takers who do not reach each item on the test"); WIGHTMAN & MULLER, *supra* note 35, at 6 (using the number of test takers who fail to answer each question on the LSAT as a measure of speededness). This measure of speededness is frequently attributed to FRANCES J. SWINEFORD, TECHNICAL MANUAL FOR USERS OF TEST ANALYSIS (ETS, Statistical Rep. 56-42, 1956).

66. WIGHTMAN & MULLER, *supra* note 35, at 6 ("In general, a test section is not considered speeded if over 90 percent complete it.").

67. *See* TWO-STATE MIXTURE MODEL, *supra* note 20, at 1 (noting that the current measure of speededness "ignores the fact that random guessing is likely to occur as time expires, especially on multiple-choice tests, such as the LSAT, that do not subtract points for wrong responses" and that commercial coaching aids and test instructions often encourage guessing).

Schnipke and Scrams noted that the current measure "probably underestimates the true amount of speededness on the test."[68] Schnipke and Scrams further observed, "[t]he primary function of the LSAT is not to measure the rate of work (speed), thus speededness is considered an ancillary variable. To know how great an effect this ancillary variable has on test scores, an accurate measure of speededness is needed."[69]

To date, the most accurate assessment of speededness appears to be the two-state mixture model developed by Schnipke and Scrams. Relying on data collected on the logical and analytical reasoning items from a nonadaptive computerized version of the GRE, which are "very similar to two of the item types (with the same name) on the current LSAT,"[70] Schnipke and Scrams matched together observations for both item accuracy and item response times. Using probability density functions, which plotted response times on the x-axis, and accuracy and proportion of test takers on the y-axis, Schnipke and Scrams observed a single-mode distribution on the first half of the test, which they hypothesized reflected "solution behavior" by the vast majority of test takers.[71] However, on the second half of the test, beginning at item 12 on a 25-question section, two modes were observed. The first mode, which the study associated with "rapid-guessing behavior,"[72] consisted of relatively fast response times and a low rate of accuracy (less than random). In contrast, the second mode, which the study associated with "solution behavior," was comprised of longer response times and a relatively high rate of accuracy.[73] An excerpt from Schnipke and Scrams' probability density functions is presented in Appendix 1.

Overall, Schnipke and Scrams observed that only 15% of test takers failed to reach the last item on the test.[74] This finding represents a relatively modest level of speededness under the traditional measure. However, the two-stage mixture model revealed that up to 53% engaged in rapid-guessing behavior, presumably in response to time constraints.[75] Further, Schnipke and Scrams found that the "underlying rapid-guessing distribution appears to be the same for all items regardless of item content, difficulty, length, or other item characteristics."[76] Other researchers have calculated that the optimal time limit on a nonadaptive GRE would permit 75% of all test takers to complete the test under power conditions, with the remaining 25%

68. *Id.*
69. *Id.*
70. *Id.*
71. *Id.*
72. *Id.*
73. *Id.* at 7 fig.1.
74. *Id.* at 14.
75. *Id.*
76. *Id.* at 15.

working under speeded conditions.[77] It is therefore not surprising that Schnipke and Scrams concluded their study with the following observation:

> Time limits are useful for administrative reasons and scheduling purposes, and if the time limits are liberal enough, there may be few negative consequences. Determining if the time limit is adequate is crucial. If nearly all test takers are able to consider fully most of the items (i.e., engage in solution behavior), the time limit has little effect on test scores, in which ease the degree of speededness is very small and can be ignored. This was not the case on the test analyzed in the present study; it was largely affected by the time limits.[78]

Because these findings are based on test items and time limitations very similar to the LSAT, it is possible that the LSAT may also contain a relatively large speededness component.[79]

Although the presence of a significant speededness component on a univariate power test will ordinarily result in a less reliable measure of the underlying construct,[80] a speeded power test may still be a relatively reliable predictor of performance on various reasoning tasks that are also performed under speeded conditions. This situation could occur if significant time constraints are imposed on both the predictor (LSAT) and the criterion measure (law school grades).[81] As a result, it is possible that the relatively high

---

77. *See* D.S. Bhola et al., *Setting an Optimum Time Limit for a Computer-Administered Test* (Paper presented at the annual meeting of the Midwestern Educational Research Association, Chicago, Oct. 1993), *cited in* TEST TAKER BEHAVIOR, *supra* note 55, at 16.

78. TWO-STATE MIXTURE MODEL, *supra* note 20, at 16.

79. This claim is corroborated by earlier research, conducted over 30 years ago, which found that LSAT scores were consistently higher when administered under unspeeded conditions. *See, e.g.*, Evans & Reilly, *supra* note 35, at 117 (finding that LSAT test takers consistently "score higher under unspeeded conditions"). Similarly, in the early 1990s, two members of the LSAC Test Development and Research Committee acknowledged that the number of wrong answers tends to increase at the end of each section, and that this pattern suggested that candidates may be guessing as they run out of time. *See* Himerfeld, *supra* note 49, at 33–34 & nn.89–90 (containing interviews with Professor Frederick Hart of the University of New Mexico School of Law and Professor David Kaye of Arizona State University College of Law).

80. *See* KLINE, *supra* note 17, at 36 ("[I]t is essential that all [psychometric] tests are univariate .... A test which measures more than one variable cannot be accurately interpreted and identical scores may not be psychologically the same.").

81. *See* L.M. Kendall, *The Effects of Varying Time Limits on Test Validity*, 24 EDUC. & PSYCHOL. MEASUREMENT 789, 790 (1964) ("If the criterion measure contains both speed and power variance, then there may be a unique predictor time limit for which the relative weightings of speed and power in the predictor scores most closely approximate the optimal weightings required for maximum validity."); Brent Baxter, *An Experimental Analysis of the Contributions of Speed and Level in an Intelligence Test*, 52 J. EDUC. PSYCH. 285, 290–91, tbl.II (1960) (obtaining measurements of accuracy and speed under both timed and untimed conditions on an intelligence test and finding that the timed version often had higher intercorrelations with aptitude tests known to be speeded while the untimed version had higher intercorrelations with other unspeeded aptitude tests).

correlation on most LSAT validity studies may be partially attributable to test-taking speed rather than reasoning ability.[82]

As noted earlier, most LSAT validity studies focus on the correlation between law school grades and the LSAT, UGPA, and a typically more robust LSAT/UGPA regression model. If the LSAT is a heavily speeded test, but actual law school exams are "power" tests largely unaffected by time constraints, we might expect the correlation coefficient of the LSAT to be too low. Yet, the absence of significant time pressure on law school exams is probably an unrealistic assumption.

Although this issue has not been a topic of systematic investigation, there appears be a general consensus among law professors that time pressure is a standard component of most law school exams. For example, in a comprehensive critique of traditional law schools exams ("Blue Books"), Professor Kissam notes that "the concept of time constraint is the most widely shared characteristic of law school exams . . . ."[83] Kissam further observes:

> Blue Book exams clearly emphasize speed in performance. With some law professors this emphasis stems from time constraints they impose on Blue Book exams for no particular reason, other than tradition. This emphasis also may stem from a conscious belief among many law professors that speed or quickness is important to many legal situations and should be measured by Blue Book exams.
>
> In major part, however, this emphasis on speed stems from a more obscure source. . . . This source is the perceived need of many schools and professors to impose grading curves with many distinctions on examination performances in individual courses. Multiple grade categories can be generated and explained to students most easily by establishing the final exam as a race and then observing the order in which contestants cross the finish line. . . . Yet, treating Blue Book exams as a race, while possibly a good test of speed, is unsound as educational practice and may even fail to measure accurately the qualities that it purports to measure.[84]

---

82. *Cf.* Hinerfeld, *supra* note 49, at 28–29. Hinerfeld argues:

> In the context of the LSAT, speededness may have neutral or even positive affect [sic] on the prediction of FYA. . . . [T]here may simply be something about the characteristics of the test that makes it a better predictor under speeded conditions. *Whether a particular level of speeding enhances the predictive validity of the LSAT is an open, empirical question. It has never been studied.*

Id.

83. Kissam, *supra* note 22, at 438; *see also* Hinerfeld, *supra* note 49, at 30 (stating in an interview that the LSAT is not designed to be a test of speed, Professor Hart of the LSAC Test Development Committee nonetheless noted: "And that's not to deny that speed is important in law school exams. It clearly is.").

84. Kissam, *supra* note 22, at 453.

Kissam's observations are corroborated by other legal academics. In several commercial study aids on law school exams written by law professors, a common theme is the importance of developing a strategy to cope with significant time pressure.[85]

A speed variable that affects both the LSAT and law school grades may also partially account for patterns in variance that have emerged on LSAT validity studies. For example, in a recent large-scale LSAT validity study, Professor Wightman found that the standardized LSAT beta weights ($ß_{LSAT}$) in the traditional LSAT and UGPA regression equation tended to slightly drop after the first year, whereas the UGPA beta weights ($ß_{UGPA}$) tended to increase, often at statistically significant levels.[86] Although Wightman found that "the LSAT score and the UGPA, in combination, were related to cumulative LGPA at approximately the same level as they were related to first-year LGPA,"[87] this result occurred because of changes in the underlying regression weights between year one and years one through three, usually by increasing the weight assigned to UGPA. An earlier LSAT validity study by Donald Powers reported similar findings, with the relative importance of UGPA as a predictor increasing during the second and third year of law school and the importance of LSAT gradually decreasing.[88] Although Powers theorized that these findings might be explained by a variety of factors, including the possibility that UGPA reflects, in part, persistence and motivation, another explanation is that less time-pressured papers and take-home exams may be more prevalent during the second and third years. In other words, the strength of correlation between the LSAT and law school grades may vary in proportion to the number of grades that are determined by speeded, in-class exams.

To accurately contextualize the relationship between the LSAT and law school performance, it is important to distinguish between two concepts of test validity: predictive validity and construct validity. Predictive validity, which is the threshold used in most LSAT validity studies, is concerned with

---

85. *See, e.g.*, WHITEBREAD, *supra* note 23, at 9 (acknowledging and defending, from his experience as a law professor at the University of Southern California, the "immense time pressure" on most law school exams); HELEN S. SHAPO & MARSHALL S. SHAPO, LAW SCHOOL WITHOUT FEAR: STRATEGIES FOR SUCCESS 171 (1996) (noting, as professors of law at Northwestern University, that "the issue is not, 'What are the best answers you can write on this exam?' Rather, it is, 'What are the best answers you can write in three hours?'"); R. MICHAEL FISCHL & JEREMY PAUL, GETTING TO MAYBE: HOW TO EXCEL ON LAW SCHOOL EXAMS 244–45 (1999) (observing, as law professors of the University of Miami and the University of Connecticut, that "[m]ost law school exams are time-pressured, some quite drastically so" and advising students to "adopt a speedy style").

86. WIGHTMAN, BEYOND FYA, *supra* note 51, at 18 tbl.5.

87. *Id.* at 2.

88. Donald E. Powers, *Long-Term Predictive and Construct Validity of Two Predictors of Law School Performance*, 74 J. EDUC. PSYCHOL. 568, 568, 570 (1982) (stating that "[t]he contribution of each measure [in predicting law school performance] was nneven across years, with the LSAT making its greatest contribution in early years and UGPA in later years").

the strength of the correlation between the predictor and the criterion measure. A consistent positive correlation is usually sufficient to establish predictive validity.[89] In contrast, construct validity, which is much more difficult to establish, requires that test designers articulate a clear theoretical basis ("construct theory")[90] for why a specific construct possessed by the test taker will be a useful predictor of some outcome in the future (such as law school grades).[91] The construct theory is then used to generate hypotheses about the test results; in turn, these hypotheses are empirically tested.[92] In essence, construct validity is a *process* that provides empirical evidence that a predictor test actually measures the individual characteristic it purports to measure.[93]

In the context of the LSAT and law school testing methods, construct validity can be coneeptualized at two distinct levels. On the first level, the LSAT is designed to measure acquired verbal reasoning and reading skills, which is a construct presumably relevant to law school success.[94] The LSAT is therefore a test that is used to predict performance on future tests—exams and writing assignments administered during law school.

The second level of construct validity focuses on the relationship between law school testing methods and the intellectual tasks and abilities that are relevant to the practice of law. Law school testing methods are designed to measure acquired legal knowledge and analytical ability, which is a construct presumably relevant to the practice of law. However, while it may be argued that lawyers work under intense time pressure, it is unclear whether the time pressure of in-class exams accurately reflects the practice environment. "Blue Book" exams require a relatively rapid analysis written within a short time period (for instance, three hours). In terms of organization, quality of analysis, and polished prose, a student response on an in-class exam is unlikely to resemble the type of work product that could

89. *See* KLINE, *supra* note 17, at 32–33 (defining and discussing predictive validity and noting the usual threshold for it to be meaningful).

90. *Id.* at 37.

91. *Id.* at 36–38 (defining and discussing construct validity and noting that "most test constructors . . . agree . . . that construct validity is the best approach to validating a test").

92. *See id.* at 37; SAMUEL MESSICK, VALIDITY OF TEST INTERPRETATION AND USE 14–19 (ETS, Research Rep. 90-11, 1990).

93. *See* KLINE, *supra* note 17, at 37 (presenting an example of how construct validity is determined); MESSICK, *supra* note 92, at 18–19 (discussing the process of construct validity); Powers, *supra* note 88, at 569 (examining LSAT and UGPA correlations for all three years of law school because "the ultimate goal [of the study] was a better understanding of the construct validity of these two measures through the *process*, described by Messick (1975), of marshaling evidence concerning relevant empirical relationships to support particular meanings of measures") (emphasis added).

94. *See* LSAC REFERENCE MANUAL, *supra* note 16, at 7 ("The LSAT is designed to measure skills that are considered essential for success in law school.... The LSAT provides a standard measure of acquired reading and verbal reasoning skills that law schools can use as one of several factors in assessing applications.").

be relied upon by a client, another lawyer, or a court.[95] The time pressure of the practice of law might be better described as managing various tasks and deadlines in order to produce legal work of sufficient quality to serve a client's interests. For example, a lawyer might feel tremendous time pressure because she has only ten days to submit a brief on a set of complex legal issues while also responding to other client demands. The time pressure is not encapsulated in discrete three-hour doses; it is systemic. The upshot of this analysis is that performance on take-home exams and papers may be a closer analogue to the practice of law.[96] Thus, a sound construct theory of law school testing may be at odds with heavy reliance on time-pressured, in-class exams.

The legal academy's emphasis on speed may also be relevant to the differential performance of white and minority students on both the LSAT[97] and on actual law school exams. For example, Wightman and Muller examined the number of unreached items on each section of the LSAT by ethnic subgroup and found a "particularly dramatic" speededness differential between African American, Puerto Rican, and Hispanic candidates.[98] Similarly, in a study that examined speededness as a potential source of bias on the LSAT, Evans and Reilly found that an experimental reading comprehension portion of the LSAT was significantly more speeded for a sample group of students from predominantly black colleges than for students from the general college population.[99]

---

95. Indeed, even if a student manages to successfully "spot" relevant issues, in-class, Blue Book exams are often poorly written, poorly organized, and unnecessarily long. *See* discussion *supra* note 23.

96. For a discussion of "speed" and the legal profession, see *infra* subpart IV(B).

97. *See generally* LSAC MINORITY DATABOOK, *supra* note 11.

98. LINDA F. WIGHTMAN & DAVID G. MULLER, COMPARISON OF LSAT PERFORMANCE AMONG SELECTED SUBGROUPS 6 (LSAC, Statistical Rep. No. 90-01, 1990).

99. Evans & Reilly, *supra* note 35, at 116, 121 figs.1a, 1b ("Figures 1a and 1b clearly indicate that the Reading Comprehension sections of the LSAT would be considered a speeded measure for fee-free candidates [who were, as a group, predominantly black] and an unspeeded measure for regular eenter candidates [who were, as a group, predominantly white]."). Despite the fact that the fee-free group achieved proportionately larger gains under unspeeded conditions, and that those results proved to be a more reliable measure of ability, the authors concluded that the reduction in speededness was "not significantly more beneficial (in terms of increasing the number of items answered correctly) to fee-free than to regular center candidates." *Id.* at 118–19. The authors also expressed concern that the fee-free sample group may not be fairly representative of black candidates generally. *Id.* at 119. A subsequent study by the same authors found that an experimental reading comprehension section of the LSAT was more speeded for black candidates generally and that, under unspeeded conditions, black females experienced larger gains *vis-à-vis* white females. *See* Franklin R. Evans & Richard R. Reilly, *The LSAT Speededness Study Revisited: Final Report* (LSAC-72-3), *in* 2 SPONSORED RESEARCH, at 191, 194–97 & fig.1, tbl.2. The authors, however, concluded that these differentials were insuffieient to warrant additional research. *Id.* at 197. Notwithstanding these conclusions, both studies by Evans and Reilly presented clear evidence that (a) *all* subgroups averaged higher scores when given more time (i.e., accuracy on the LSAT was partially a function of time), and (b) black test takers were significantly less likely to answer all items on the test. Furthermore, an important limitation of the two studies by Evans and Reilly is

The hypothesis examined in this study is whether both the LSAT and traditional law school exams are affected by a significant speededness component. Yet, if the hypothesis is true, then a differential in testing speed among ethnic subgroups may be hard to discern from LSAT validity studies. In general, research on the LSAT has often presented some rather puzzling results regarding minority students. For example, several studies have documented the tendency of the LSAT/UGPA regression model to "over-predict"[100] the performance of minority students during the first year of law school.[101] Similarly, other studies have found that the LSAT tends to over-predict the performance of minority students during all three years of law school, though this tendency is reduced somewhat when UGPA is included in the regression equation.[102] The problem of over-prediction should be compared to the findings of a recent article by William Kidder.[103] After controlling for undergraduate school, major, and UGPA, Kidder found that minority students in the Boalt Hall applicant pool during the late 1990s scored on average between 3.6 and 9.1 points lower on the LSAT than their white counterparts.[104]

Different rates of test-taking speed among minority subgroups may partially explain these patterns. For example, if minority students tend to be more affected by a speededness component *and* law school exams are even more speeded than the LSAT, then over-prediction may result. Similarly, if most undergraduate testing methods are unspeeded, then differences in test-taking speed among minority subgroups may result in systematically lower

---

that they focused only on the reading comprehension section of the LSAT. However, as discussed below, some commentators view the analytical reasoning section as the most time-pressured section of the exam. *See infra* note 110 and accompanying text.

100. Over-prediction occurs when the LSAT predicts a better performance than actually observed within a particular subgroup. WIGHTMAN & MULLER, *supra* note 98, at 1.

101. *See id.* at 1 (examining data from 54 law schools and noting that the LSAT/UGPA model "tends to overpredict performance for minority students"); *see also* Donald E. Powers, *Comparing Predictions of Law School Performance for Black, Chicano, and White Law Students* (LSAC-77-3), *in* 3 REPORTS OF LSAC SPONSORED RESEARCH, 1975–1977, at 721, 744–45 (1977) (examining data from 31 law schools and discussing the tendency of the LSAT/UGPA model to over-predict for black and Chicano students); W.B. Schrader & Barbara Pitcher, *Prediction of Law School Grades for Mexican American and Black American Students* (LSAC-74-8), *in* 2 SPONSORED RESEARCH, *supra* note 35, at 715, 717 (reviewing correlation data from seven law schools and noting the tendency of the LSAT/UGPA model to over-predict performance of black and Mexican students).

102. *See* WIGHTMAN, *supra* note 51, at 2 (reporting that "over-prediction [for nonwhite students] was greater when [an] LSAT score was used alone than when it was used in combination with UGPA"); Donald E. Powers, *Predicting Law School Grades for Minority and Nonminority Students: Beyond the First-Year Average Grades* (LSAC-81-1), *in* 4 SPONSORED RESEARCH, at 261, 281–84 & tbl. (finding similar results based on a sample of 23 law schools).

103. William C. Kidder, Comment, *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 (2001).

104. *Id.* at 1079–80 & tbl.3 (finding gaps of 9.1 for African Americans, 7.0 for Latinos, and 3.6 for Asian Pacific Americans).

LSAT scores even though the test takers had virtually identical UGPAs from the same college in the same major. In terms of the present study, this explanation of minority subgroup performance would be supported if minority students, when compared to white students, perform significantly better on less time-pressured testing methods, such as take-home exams and papers, versus in-class exams.[105]

Finally, a large speededness component on the LSAT may also be relevant to research on LSAT subscore discrepancies. Subscore discrepancies are performance differentials between the logical reasoning, analytical reasoning, and reading comprehension sections of the LSAT.[106] Because subscores generally show a strong intercorrelation, discrepancies are useful in interpreting total test score results.[107] Based on a sample of 39,530 examinees from an actual LSAT administration, Lawrence Stricker found that subscore discrepancies did not significantly track ethnicity, gender, or test familiarity.[108] However, Stricker also found that approximately one-third of examinees had statistically significant differentials and that this tendency was greatest among examinees "who had high total scores on the LSAT or were older, primarily reflecting the test takers' deviantly poor performance on the Analytical Reasoning subtest."[109]

The finding that older students tend to do worse on the analytical reasoning portion of the LSAT may be related to differences in test-taking speed. Although the methodologies of test preparation companies are far from scientific, one of the leading LSAT study aids singles out analytical reasoning as the most time-compressed on the test, suggesting that many test takers will have time to attempt only three out of four groups of questions, with the remaining group (roughly six questions) completed through rapid guessing.[110] This observation should be compared to findings within the psychometric literature cautioning that the variable of age must be controlled in correlation studies because test-taking speed diminishes as one ages.[111]

---

105. These comparisons are made *infra* section IV(C)(4).

106. *See* LAWRENCE J. STRICKER, DISCREPANT LSAT SUBSCORES 1 (LSAC, Research Rep. No. 93-01, 1993).

107. *Id.*

108. *Id.* at 1–2.

109. *Id.*

110. *See* ROBINSON & TALLIA, *supra* note 50, at 66. The authors note:

> The LSAT writers have put four games in the [analytical reasoning] section to scare you into trying to work too fast, thereby cutting down on your accuracy. If you're able to get to three games in the section, and get all the questions on those three games correct, chances are you'll receive a very good LSAT score.

Id.

111. *See generally* TIMOTHY A. SALTHOUSE, THE HANDBOOK OF AGING AND COGNITION (2000); TIMOTHY A. SALTHOUSE, A THEORY OF AGING AND COGNITION (1985) (both summarizing the effect aging has on cognitive abilities). *See also* ARTHUR R. JENSEN, THE G FACTOR: THE

In terms of the present study, the variable of age may be relevant to explain differential performance on testing methods with shorter versus longer time limits. Similarly, because Stricker found that students with generally high LSAT scores are more prone to large subscore discrepancies on the analytical reasoning portion of the test, it may be possible that the national law school sample, which contains a very large proportion of high LSATs,[112] may be relatively homogeneous in terms of reasoning ability but relatively heterogeneous in terms of student test-taking speed. Thus, within this population, we might expect larger changes in LSAT correlation coefficients when the LSAT is correlated to testing methods with short versus long time limitations.

III. Sample and Methodology

A. Sample

The data for this study was obtained from two law schools with different student populations.[113] For the purposes of this study, the "national law school sample" refers to the data set drawn from the law school with relatively high UGPA and LSAT scores. In contrast, the "regional law school sample" denotes the data set from the law school with mid-range UGPA and LSAT scores. The primary reason for including two law schools in this study is to provide some initial indication of whether any speededness effect is more likely to operate within a specific range of UGPA or LSAT scores.

The national law school sample contains 379 students who recently graduated from the same elite national law school. It is important to emphasize that this sample represents a relatively narrow range of the total population of law school graduates who attended a U.S. law school during the same time period. For example, 90% of all LSAT scores in this sample are 161 or higher. Similarly, 90% of all undergraduate grade point averages are 3.25 or higher. Therefore, the results of this sample cannot be generalized to other law schools with different student populations. Table 4 contains a breakdown of the entering credentials by gender, ethnicity, and transfer status.

SCIENCE OF MENTAL ABILITY 225 (1998) (noting the importance of controlling for age because "test-taking speed . . . change[s] for the 'worse' with increasing age beyond early adulthood").

112. *See infra* tbl.4.

113. The data utilized in this study was obtained on a condition of confidentiality for the two participating law schools. The agreement for grant funding also required that the two law schools remain anonymous. However, because the tables in this subpart summarize each sample population in great detail, the identity of each law school is irrelevant.

Table 4: Summary of National Law School Sample by LSAT & UGPA

Population		*Size*		*LSAT Score by Percentile*			*UGPA by Percentile*		
		n	%	25th	50th	75th	25th	50th	75th
Sex	*Male*	233	61.5	165	168	171	3.42	3.60	3.77
	Female	146	38.5	164	167	170	3.48	3.62	3.78
Ethnicity	*White*	264	69.7	165	168	171	3.47	3.61	3.78
	Asian/Pac. Islander	34	9.0	164	168	170	3.51	3.69	3.79
	Black	13	3.4	160	161	166	3.33	3.51	3.67
	Hispanic†	18	4.7	161	165	169	3.20	3.36	3.59
	Undeclared	50	13.2	165	169	171	3.46	3.63	3.81
Transfer	*Entered as 1L*	343	90.5	165	168	171	3.49	3.63	3.79
	Transferred In	36	9.5	157	161	166	2.99	3.29	3.47
All Students In Sample		379	100	164	168	171	3.45	3.61	3.77

† This category includes students who identify themselves as Mexican, Puerto Rican, or Other Hispanic.

The restriction of range[114] in the national law school sample, as evidenced by the large proportion of high LSAT scores, has the potential of distorting the relative importance of the LSAT and UGPA as predictors of law school performance. In general, the LSAT becomes a better predictor of law school performance (i.e., it has a higher correlation with law school grades) as the range of LSAT scores (and thus ability) becomes wider. In

114. Range restriction refers to the limitation on the predictor variable to predict future academic performance when the sample group is composed of very similar students in terms of LSAT scores, as in the national law school sample. If the range of scores was broader, the predictive validity of the LSAT would increase significantly. *See* Wightman, *supra* note 32, at 33 (describing the results of an empirical study conducted by a former LSAC psychometrician and explaining that "restriction of range" and "the validity coefficients reported for the LSAT tend to be underestimates" because the range of enrolled students tends to be much narrower than the total population who applied).

contrast, when the range of LSAT scores is relatively restricted, as in this population, the LSAT will become a less accurate predictor of law school performance. Although UGPA may be similarly affected by range restriction, a visual inspection of LSAT and UGPA percentiles as reported in *U.S. News & World Report* suggests that the range of overlap of LSAT scores between law schools is generally smaller than the range of overlap of UGPA.

The potential effect of LSAT range restriction on the national law school sample can be observed when we compare the transfer student population to the sample as a whole. Among the entire student population, which has a relatively narrow LSAT range,[115] the correlation between LSAT score and final law school grade point average (LGPA) is a relatively modest 0.194. However, within the transfer student population, which had a lower median LSAT score and a much broader range, the correlation between LSAT scores and LGPA is 0.476. Similarly, while UGPA was a more robust predictor of LGPA for the national law school sample as a whole,[116] UGPA had no statistically significant predictive force among the transfer students.[117]

The regional law school sample is approximately 50% larger than the sample drawn from the national law school.[118] The regional law school sample contains 604 students who recently graduated from the same regional law school.[119] Of these 604 students, approximately 153 entered law school as part-time evening students.[120] Although these students had, on average, slightly lower UGPA and LSAT scores than the day students, a substantial minority eventually entered the full-time day program.[121] Moreover, after the first year of law school, it was not uncommon for day students to take night classes and for night students to take day classes. As a result, the regional law school sample has limited value for analyzing distinctions between these two groups. Table 5 contains a breakdown of the entering credentials by gender, ethnicity, transfer status, and whether the student initially enrolled in the day or evening division.

---

115. Although the numerical range of LSAT scores in the national sample is relatively broad (over 30 points), the vast majority of students were in the 90th percentile of all LSAT test-takers. *See infra* tbl.4.

116. *See infra* tbls.9 &16.

117. The poor predictive ability of UGPA for transfer students is not surprising. Transfer students are typically admitted based on their strong first-year performance at another law school; thus, for this population in partieular, numerical criteria based on undcrgraduate performance may be more likely to underpredict student performance.

118. *Compare supra* tbl.4 *with infra* tbl.5.

119. *See infra* tbl.5.

120. *Id.*

121. *Id.*

Table 5: Summary of Regional Law School Sample by LSAT & UGPA

Population		*Size*		LSAT Score by Percentile			UGPA by Percentile		
		n	%	25th	50th	75th	25th	50th	75th
Sex	*Male*	301	49.8	150	153	157	2.79	3.08	3.32
	Female	303	50.2	149	153	156	2.86	3.11	3.42
Ethnicity	*White*	486	80.5	150	154	157	2.85	3.11	3.38
	Asian/Pac. Islander	31	5.1	151	153	155	2.70	2.93	3.14
	Black	28	4.6	146	149	154	2.66	2.96	3.41
	Hispanic†	24	4.0	146	149	153	2.79	2.99	3.29
	Undeclared	35	5.8	149	152	155	2.86	3.11	3.36
Transfer	*Entered as IL*	587	97.2	150	153	157	2.83	3.10	3.38
	Transferred In	17	2.8	147	151	157	2.70	3.05	3.28
Day / Evening	*Entered as full-time day*	451	74.7	150	154	157	2.88	3.11	3.38
	Entered as full time eve	153	25.3	147	152	156	2.72	3.03	3.36
All Students In Sample		604	100	150	153	157	2.83	3.10	3.37

† This category includes students who identify themselves as Mexican, Puerto Rican or Other Hispanic.

In the process of assembling the data set, we coded for a variety of variables that may be useful to this project or other future research. In addition to LSAT, UGPA, sex, ethnicity, and transfer status, both the national and regional law school samples include variables for (a) age upon entry to law school, (b) undergraduate major (four categories), (c) whether the student attended an elite or nonelite undergraduate institution,[122] and (d) whether the student earned a graduate degree before entering law school. A summary of these categories is presented in Table 6.

Table 6: Breakdown by Age, Undergraduate Major, Undergraduate School, and Graduate Degrees

Characteristic	*Breakdown*	*National Law School*	*Regional Law School*
Age	Mean	24.4	25.7
	Median	23.5	24.0
	Range	20–42	20–58
Undergraduate Major	Liberal Arts	35.9%	31.6%
	Social Science	46.9%	41.4%
	Business/Accounting	4.7%	11.4%
	Science/Engineering	12.6%	14.2%
Undergraduate School	Elite	56.8%	11.6%
	Non-Elite	43.2%	88.4%
Graduate Degree	Earned Before Law School	5.5%	8.1%
	No Graduate Degree	94.5%	91.9%

Finally, the data sets for both the national and regional law school samples are very detailed. The national law school sample contains approximately 17,700 individual grades from actual law school courses, seminars, clinics, and independent research projects. Of these 17,700 individual grades, approximately 11,300 were given to students in the sample group. The regional law school sample contains approximately 26,000 individual grades; of these, approximately 15,300 were given to students in the sample group. In both the national and regional law school samples, the nonsample grades correspond to students who took classes during the relevant time period but were either L.L.M. students or J.D. students from a different graduating class. The larger data sets were needed to accurately determine the mean, median, number of students, and grade distribution for

122. Using charts in *America's Best Colleges*, U.S. NEWS & WORLD REP. (2003), the author labeled 40 colleges and universities as "elite" based on SAT scores and acceptance rates. All other undergraduate institutions were labeled non "elite." Students were then coded accordingly (list of 40 schools on file with author).

cach course given during the sample period. Thereafter, relevant correlation and regression calculations were derived from the 11,300 grades in the national law school sample and the 15,300 grades in the regional law school sample. A summary of average student GPAs by testing method is presented in Table 7. Histograms for average student GPAs by testing method are also contained in Appendices 2 and 3.

Table 7: Summary Statistics of Average Grades by Testing Method, National & Regional Law School Samples

Statistic on GPA by Testing Method		*GPA (In-Class)*	*GPA (Take-Home)*	*GPA (Papers)*
National	*N*	379	377	379
	Mean	3.08	3.18	3.42
	Median	3.06	3.18	3.44
	Std. Deviation	.269	.293	.230
	Minimum	2.20	1.70	2.54
	Maximum	3.87	4.10	4.00
Regional	*N*	603	482	601
	Mean	3.00	3.14	3.24
	Median	2.99	3.18	3.25
	Std. Deviation	.395	.504	.373
	Minimum	1.89	1.00	1.83
	Maximum	3.93	4.30	4.00

B. Analysis Methods

The central hypothesis of this project is that test-taking speed is a variable that affects performance on both the LSAT and law school exams. However, because this variable cannot be directly and unobtrusively measured under actual test conditions, this project relies upon a two-phase methodology to indirectly assess the possible effect of test-taking speed. In Phase 1, student performance was disaggregated into three discrete averages: (1) in-class exams (~3 hours), (2) take-home exams (~8 to 24 hours), and (3) assigned papers. LSAT and UGPA correlation coefficients were then generated for each testing method. Because time constraints on the LSAT are most similar to in-class exams, the hypothesis on test-taking speed was assumed to be supported if the LSAT was a significantly better predictor of

performance on in-class exams than less time-compressed, take-home exams and papers.[123] In other words, the following relationships should be present:

$$(R_{\text{in-class}} > R_{\text{take-home}} > R_{\text{paper}}) \quad \text{or} \quad (R_{\text{in-class}} > R_{\text{take-home \& paper}})$$

A "z" test was then used to determine whether the divergence between the correlation coefficients was statistically significant.

In Phase 2, the objective was to create regression equations that predicted performance on each testing method. These equations were then used to create a model of law school performance that effectively included testing method as a variable. The equations for predicting performance on each testing method can be summarized as follows:

$$GPA_{\text{in-class}} = k_{\text{in-class}} + \beta_1(LSAT) + \beta_2(UGPA)$$

$$GPA_{\text{take-home}} = k_{\text{take-home}} + \beta_3(LSAT) + \beta_4(UGPA)$$

$$GPA_{\text{paper}} = k_{\text{paper}} + \beta_5(LSAT) + \beta_6(UGPA)$$

Cumulative law school average (GPA_{law}) was then predicted based on the proportion of law school credits (P), specific to each student, that is attributable to each testing method. It should be noted that a small proportion of grades in the sample were given in clinical classes (national, 2.5%; regional, 5.2%) or in classes in which the testing method was unknown (national, 1.0%; regional, 0.2%). Therefore, regression equations were generated for these categories as well. For all students in the sample, $P_{\text{in-class}} + P_{\text{take-home}} + P_{\text{paper}} + P_{\text{clinic}} + P_{\text{unk.}}$ equals 1.0. The disaggregated model for predicting law school performance takes the following form:

123. In both sample groups, grades on papers and take-home exams tended to be higher, see tbl.7, primarily because these testing methods were more common in small, upper-level classes, which were not subject to mandatory grading curves. However, a higher grade distribution would not, by itself, produce significantly lower correlation between LSAT (or UGPA) and grades. This is because correlation coefficients reflect the strength of the relationship between the LSAT (or UGPA) and the *ordering* of student performance on each testing method (i.e., performance *vis-à-vis* fellow students). Even if GPAs for take-home exams and papers are higher than GPAs for in-class exams, LSAT (or UGPA) correlation coefficients for each testing method will be virtually identical if the ordinal ranking between testing methods remain the same (i.e., the person who does best on in-class exams also does the best on papers and take-homes, the person who does worst on in-class exams also does the worst on papers and take-homes, etc.). In this sample, GPAs for each testing method also had fairly broad distributions that conformed to a normal curve. *See* Appendices 2 and 3 and tbl.7. In other words, within each testing method, there were significant variations in performance that were not bunched at the top of the grading scale. Finally, if differences in grading curves produced substantial variations in the correlation coefficients of each testing method, the effect on LSAT and UGPA correlation coefficients should be similar. Yet, on all testing methods in both samples, the correlation coefficients for UGPA were much more stable than the correlation coefficients for the LSAT. *See infra* subpart IV(A). This pattern was also reflected in the LSAT and UGPA standardized regression weights that were generated for each testing method. *See infra* subpart IV(B).

$$GPA_{law} = P_{in\text{-}class}(GPA_{in\text{-}class}) + P_{take\text{-}home}(GPA_{take\text{-}home}) + P_{paper}(GPA_{paper}) + P_{clinic}(GPA_{clinic}) + P_{unk.}(GPA_{unk.})$$

Next, the predictive power of the disaggregated model was compared to the predictive power of the traditional LSAT/UGPA regression model. For reasons of clarity, the traditional LSAT/UGPA model is referred to as the "aggregated model." Correlation coefficients for both the disaggregated model (R_D) and aggregated model (R_A) were calculated based on actual GPA_{law} and predicted GPA_{law}. If the difference between R_D and R_A is both positive ($R_D - R_A > 0$) and statistically significant, it is presumed that testing method is a variable within that sample group. The variance in law school performance attributable to testing method is therefore the difference between $(R_D)^2$ and $(R_A)^2$. However, it should be noted that even if there are large differences in how the LSAT and UGPA predict each testing method, the disaggregated model is inevitably limited by the proportion of testing methods in each sample. In other words, if one testing method dominates the sample, there may be very little to disaggregate.

Finally, when interpreting results from Phases 1 and 2, it is important to distinguish between two variables: testing method and test-taking speed. Specifically, this project is designed to measure how well the LSAT and UGPA predict first-year and overall law school performance on different testing methods defined by time allowed (in-class exams, take-home exams, and papers). Thus, in a narrow sense, testing method is the variable that is being directly isolated. In contrast, the test-taking speed hypothesis predicts that the LSAT will be a robust predictor of in-class exams and a less accurate predictor of take-home exams and papers. In the case of in-class exams, this divergence might occur because both the predictor variable (LSAT) and criterion variable (law school grades) are affected by a relatively large speededness component. In turn, when test-taking speed becomes less relevant to performance (presumably on take-homes and papers), the LSAT loses some of its predictive ability. Insofar as this pattern emerges from the sample data, the test-taking speed hypothesis is indirectly supported.[124]

Because the data from the national and regional law schools generally produced positive results during Phases I and 2, this project also included some additional analysis designed to corroborate or rebut the presence of the test-taking speed variable. The methods relevant to each inquiry are presented with the results in subpart IV(C).

---

124. The test-taking speed hypothesis could be directly tested by (a) giving each student a test that measured test-taking speed, and (b) correlating these results with LSAT scores and actual law school performance. A similar method was used by Baxter in his study of the effects of speed on intelligence tests. *See* Baxter, *supra* note 81, at 286. However, this direct method would be difficult to implement for a relatively large sample size attending the same law school during the same period.

IV. Results

The results of this study are organized into three sections. Subpart A presents the results of Phase 1. Subpart B presents the results of Phase 2. Finally, subpart C includes the results of various subsidiary analyses designed to corroborate or rebut the presence of a test-taking speed variable on both the LSAT and law school exams. Subpart C also contains subgroup analyses related to age and ethnicity.

A. Phase 1: LSAT and UGPA Correlation Coefficients by Testing Method

In Phase 1, student grades were disaggregated according to three distinct testing methods: (1) in-class exams, (2) take-home exams, and (3) assigned papers.[125] Although both law schools utilized these testing methods, it should be emphasized that the national and regional law schools reflected distinctively different testing norms that limit the comparability of the two data sets. For example, in-class exams at the national law school tended to be conducted under more stringent time limits. In-class exams at the regional law school also frequently included a mixture of multiple choice and essay questions. In contrast, this format was relatively rare at the national law school. Table 8 is a comparison of in-class exams by law school.

Table 8: Characteristics of In-Class Exams by Law School

Characteristic	*National Law School*	*Regional Law School*
Average Time	2 hours, 52 minutes	3 hours, 15 minutes
Median Time	3 hours	3 hours
Standard Deviation	21 minutes	35 minutes
Range	1.5 to 4 hours	1.5 to 5 hours
Percentage With Some or All Multiple Choice or True/False	5.1%	44.9%

125. In both the national and regional law school sample, the disaggregation process included clinical courses and courses in which the testing method could not be reliably determined ("unknown"). Although these categories were used to calculate law school averages, they are not directly relevant to the hypothesis on test-taking speed.

The national and regional law schools also had distinctively different norms regarding take-home exams. For example, the vast majority of take-home exams at the national law school (95.8%) were given over an 8-hour time period, usually with a specific word or page limit.[126] In contrast, take-home exams at the regional law school were usually administered over a 24-hour time period. Although we were unable to obtain accurate time limits for all take-home exams in the regional law school, some faculty members recalled time limits that extended over more than one week of the exam period. In fact, one faculty member stated that his take-home exams were very similar to paper assignments, with the exception that students were responding to specific test questions. Finally, it should be noted that take-home exams were much more common at the national law school. Only 6.2% of all credit hours at the regional law school were given in courses using take-home exams compared to 15.2% at the national law school.[127] The relatively low number of take-home exams at the regional law school also means that the idiosyncrasies of a relatively small number of faculty members (who wrote and graded these exams) exert a stronger influence on the sample and make the results less reliable.[128]

Attention should be paid to the above limitations when making comparisons between the two law schools. That said, the disaggregated LSAT and UGPA correlation coefficients for the national and regional law schools are presented in Tables 9 and 10 respectively.

126. It is unclear whether the decision to give a take-home exam rather than an in-class exam is, more often than not, based on a belief that one method is a better measure of student performance. According to the registrar at the national law school, faculty members' primary rationale for giving take-home exams was either: (a) to ensure typewritten responses, or (b) to facilitate student answers that were better written and organized. Obviously, both of these outcomes make the exams less onerous to grade.

127. *Compare* tbl.17 *with* tbl.20.

128. This might be referred to as an "instructor" variable, which will be insignificant in a large sample.

Table 9: Correlation Coefficients for LSAT & UGPA with Grades by Testing Method, National Law School Sample, Year 1 & Years 1-3

Testing Method	*Year 1*			*Years 1–3*		
	R_{LSAT}	R_{UGPA}	n	R_{LSAT}	R_{UGPA}	n
In-Class Exams	.236**	.291**	2,293	.265**	.296**	6,929
Take-Home Exams (~ 8 Hrs.)	.107	.270**	416	.148**	.231**	1,725
Assigned Papers (Untimed)	.006	.200**	354	.057	.192**	2,261
All Methods	.230**	.323**	3,063	.194**	.262**	10,915

* Significant at p = .05, ** Significant at p = .01

Table 10: Correlation Coefficients for LSAT & UGPA with Grades by Testing Method, Regional Law School Sample, Year 1 & Years 1-3

Testing Method	*Year 1*			*Years 1–3*		
	R_{LSAT}	R_{UGPA}	n	R_{LSAT}	R_{UGPA}	n
In-Class Exams	.480**	.269**	3,249	.453**	.323**	10,232
Take-Home Exams (~ 24 Hrs.)	.310**	.102	153	.182**	.191**	996
Assigned Papers (Untimed)	.347**	.324**	1,138	.285**	.333**	2,837
All Methods	.488**	.297**	4,540	.446**	.347**	14,065

* Significant at p = .05, ** Significant at p = .01

When comparing Tables 9 and 10, the most striking difference is that the predictive power of the LSAT varies significantly between the two samples. In the regional law school sample, the LSAT is clearly a more robust predictor, with a 0.488 correlation on all methods during year 1 and a 0.446 correlation in years 1-3. Moreover, during both of these time periods, the LSAT was a stronger predictor than UGPA. The LSAT and UGPA correlation coefficients for the regional law school sample are very similar to those reported in recent validity studies.[129] In contrast, the LSAT correlations for all methods in the national law school sample are relatively

129. For discussions that summarize the correlation coefficients for the LSAT, UGPA, and the LSAT/UGPA regression model in large multi-school samples, see, for example, WIGHTMAN, *supra* note 51, at 16 tbl.4 and ANTHONY ET AL., *supra* note 51, at 10–11 tbls.3–4.

weak. Thus, the UGPA emerged as the stronger predictor on all testing methods during both year 1 and years 1-3 of the national law school sample. This result is contrary to most LSAT validity studies.[130] However, as noted earlier, the lower correlations are probably the result of LSAT range restriction.[131] Yet, despite the different predictive power of the LSAT in each sample, it is noteworthy that the UGPA correlation coefficients for the national and regional law school are fairly similar.

1. Phase 1 Results, National Law School Sample.—In the national law school sample, the predicted pattern of divergence of LSAT correlations is both present and statistically significant. As shown in Table 11, the LSAT correlation coefficient is highest for in-class exams and progressively declines as we move to take-home exams and papers. For year 1 and years 1-3, the following relationship is present: ($R_{in\text{-}class} > R_{take\text{-}home} > R_{paper}$).

Table 11: Comparison of LSAT Correlation Coefficients, National Law School Sample

	$R_{in\text{-}class}$	$R_{take\text{-}home}$	R_{paper}
Year 1	.236**	.107	.006
Years 1–3	.265**	.148**	.057

* Significant at p = .05, ** Significant at p = .01.

Using a "z" test,[132] the LSAT correlation coefficients for in-class exams and take-home exams are statistically different during year 1 and years 1-3 (p = .01). However, the LSAT correlation coefficients for take-home exams and papers are statistically different only during years 1-3 (p = .01).

In contrast to LSAT scores, UGPA was significantly correlated with law school performance on all three testing methods. One of the hypotheses of the proposal phase of this study was that UGPA would have more predictive power on papers and take-home exams and less predictive power on in-class exams ($R_{in\text{-}class} < R_{take\text{-}home} < R_{paper}$). This relationship was based on the theory that most undergraduate testing methods are less time pressured than a traditional law school "Blue Book" exam. However, within the national law school sample, the exact opposite pattern emerged ($R_{in\text{-}class} > R_{take\text{-}home} > R_{paper}$). These values are compared in Table 12.

130. *See generally supra* note 53.

131. *See supra* note 114 and accompanying text.

132. *See* ROBERT G.D. STEEL, JAMES M. TORRIE, & DAVID A. DICKEY, PRINCIPLES AND PROCEDURES OF STATISTICS: A BIOMEDICAL APPROACH 297–99 (1997) (explaining how the "z" test is used to determine whether two or more non-zero correlation coefficients are statistically different from one another).

Table 12: Comparison of UGPA Correlation Coefficients, National Law School Sample

	$R_{in\text{-}class}$	$R_{take\text{-}home}$	R_{paper}
Year 1	.290**	.270**	.200**
Years 1–3	.296**	.231**	.192**

* Significant at p = .05, ** Significant at p = .01.

Using a "z" test, the UGPA correlation coefficients for in-class exams and take-home exams are statistically different only during years 1-3 (p = .01). The correlation coefficients for in-class exams and papers are statistically different for year 1 (p = .05) and years 1-3 (p = .01). The correlation coefficients for take-home exams and papers were not statistically different for any time period.

2. *Phase 1 Results, Regional Law School Sample.*—In the regional law school sample, performance on in-class exams once again produced the strongest correlations with the LSAT scores. In contrast to the national law school sample, grade averages for papers rather than take-home exams produced the second highest correlations during both year 1 and years 1-3, suggesting that the variance between testing methods was something more than a pure time variable. However, the LSAT correlation coefficient still fit one of the two predicted relationships ($R_{in\text{-}class} > R_{take\text{-}home\ \&\ paper}$). A comparison of LSAT correlation coefficients is set forth in Table 13.

Table 13: Comparison of LSAT Correlation Coefficients, Regional Law School Sample

	$R_{in\text{-}class}$	$R_{take\text{-}home}$	R_{paper}
Year 1	.480**	.310**	.347**
Years 1–3	.453**	.182**	.285**

* Significant at p = .05, ** Significant at p = .01.

Using a "z" test, the correlation coefficients for in-class exams were statistically different than take-home exams and papers for both year 1 and years 1-3 (p = .01). The difference between take-home exams and papers was only statistically significant for years 1-3 (p = .01).

In the regional law school sample, performance on papers produced the highest correlation with UGPA during both year 1 and years 1-3. As shown in Table 14, the lowest correlations were consistently obtained on take-home exams.[133]

Table 14: Comparison of UGPA Correlation Coefficients, Regional Law School Sample

	$R_{in\text{-}class}$	$R_{take\text{-}home}$	R_{paper}
Year 1	.269**	.102	.324**
Years 1–3	.323**	.191**	.333**

* Significant at p = .05, ** Significant at p = .01.

Using a "z" test, the UGPA correlation coefficients for in-class exams and take-home exams are statistically different for year 1 (p = .05) and years 1–3 (p = .01). The correlation coefficients for in-class exams and papers are also statistically different for year 1 (p = .05) but statistically the same for years 1–3. Finally, the correlation coefficients for take-home exams and papers are statistically different for year 1 and years 1–3 (p = .01).

3. Comparison of National and Regional Sample.—When comparing the two samples, one of the most striking findings is the relatively strong correlation between the LSAT and GPA_{paper} at the regional law school versus no relationship at the national law school. However, upon closer inspection of the regional law school sample, it was discovered that the LSAT correlations for first-year legal writing and research, which was the only class graded by paper in year 1, were significantly higher than the correlation for other courses graded by paper. Moreover, because students at the regional law school generally earned a relatively small number of credits by writing papers, the grade in first-year legal writing exerted a strong effect on the correlation coefficient for all papers during years 1–3. Table 15 summarizes the LSAT correlation coefficients after disaggregating courses graded by paper.

133. As noted earlier, the relatively small number of take-home exams in the regional law school sample (996) presents the possibility that the variance between take-home and other testing methods may be partially explained by the idiosyncrasies of the instructors who wrote and graded the exams rather than the attributes that are common to the take-home testing method. *See supra* note 128 and accompanying text. As the number of grades within a testing method increases, we become more confident that the results are not unduly influenced by an "instructor variable." Thus, the findings of statistical differences between testing methods need to be interpreted with caution.

Table 15: LSAT Correlation Coefficients for Paper Courses, National and Regional Law Schools

Paper Course	*National Law School*	*Regional Law School*
First-Year Required Legal Writing	.036	.347**
Second-Year Required Legal Writing	n/a	.106*
Seminars & Elective Courses	.053	.124**

* Significant at p = .05, ** Significant at p = .01.

In the national law school sample, the LSAT was uncorrelated with all paper grades, including the required first-year legal writing course. In contrast, the regional law school sample produces a fairly high correlation with first-year required legal writing (.347), which trailed off significantly during the second and third years of law school (.124 for seminar and elective courses, .106 for second-year legal writing). However, because first-year legal writing accounted for 44% of all credits earned in paper classes at the regional law school,[134] the overall correlation coefficient for papers remained relatively high (.285).

The difference between the regional and national law school samples in terms of LSAT correlation coefficients for papers may be partially explained by a combination of range restriction and the unique grading criteria for legal writing.[135] Specifically, first-year legal writing courses generally require students to identify and analyze legal issues from a hypothetical case. Thus, like a traditional law school exam, students are awarded credit for "spotting" issues.[136] However, unlike traditional law school exams, a substantial amount of credit is also awarded for clear, well-organized prose akin to a brief filed with a court. Thus, at the regional law school, where the student body has a broader range of academic ability than at a national law school, a larger percentage of students may fail to correctly identify and analyze relevant issues, leading to a higher correlation with the LSAT. In contrast, it is possible that students at the national law school are more likely to identify and analyze relevant issues on their first-year legal writing assignments, thus

134. *See infra* tbl.20.

135. My analysis of the grading criteria of legal writing courses versus substantive courses benefited enormously from conversations with Joseph Morrissey, who has taught both legal writing and regular law school courses that included a final exam.

136. *See* WHITEBREAD, *supra* note 23, at 7 (asserting that from the perspective of a law professor who writes a prominent study aid for law school exams, "the most important task [on a law school exam] is spotting the legal issues").

shifting more of the grading emphasis to writing style and organization. Under this framework, when the grading criteria is primarily quality of writing and organization (such as in seminars and in elective paper courses in the second and third years of law school), the LSAT has little or no predictive power in either sample.

4. Summary of Phase 1 Results.—At least three significant findings emerge from Phase 1. First, within both the national and regional samples, testing method appears to be a variable that affects the predictive power of both the LSAT and UGPA. Second, the LSAT's predictive ability is strongest on in-class exams, which is the most common testing method at most U.S. law schools.[137] Third, the predictive power of the LSAT declines at statistically significant levels as we move from in-class exams to take-home exams and assigned papers. Because these results are consistent with the hypothesis that test-taking speed is a variable affecting both the LSAT and law school performance, the project moves to Phase 2.

B. Phase 2: Constructing and Testing the Disaggregated Model

In Phase 2, the objective is to create a model of law school performance that includes law school testing method as a variable. This model is referred to as the disaggregated model. However, to establish a clear point of comparison, we initially calculated the traditional LSAT/UGPA regression model, which is based on combined grades (the aggregated model). We then generated separate regression equations for in-class exams, take-home exams, paper courses, clinical courses, and courses in which the method of grading could not be accurately determined. The proportion of each testing method for each student is then used to predict individual performance. Finally, to assess the role of testing method in law school performance, the correlation between actual and predicted law school performance for the disaggregated model (R_D) is compared to the same value for the aggregated model (R_A).

1. Phase 2 Results, National Law School.—Table 16 shows the number of students, the standardized regression weights (β), and correlation coefficients for each testing method within the national law school sample. The regression models forced in both the LSAT and UGPA variables even when a stepwise forward regression would have excluded one of them due to lack of statistical significance.

137. *See* sources cited *supra* note 56.

Table 16: Standardized Beta Weights and Correlation Coefficients by Testing Method, National Law School Sample

Testing Method	*Year 1*				*Years 1–3*			
	n	β_{LSAT}	β_{UGPA}	R	n	β_{LSAT}	β_{UGPA}	R
Aggregated Model	341	.243**	.332**	.404**	379	.161**	.240**	.307**
In-Class Exams	341	.248**	.300**	.381**	379	.229**	.264**	.373**
Take-Home Exams	274	.114	.272**	.293**	377	.118*	.215**	.259**
Papers	341	.014	.200**	.200**	379	.031	.192**	.195**

* Significant at p = .05, ** Significant at p = .01.

As reflected in Table 16, one of the most notable results in the national law school sample is that the standardized regression weights (β_{LSAT} and β_{UGPA}) for the aggregated model decline significantly from year 1 to years 1–3 (0.404 to 0.307) while the correlation coefficients for the individual testing methods decline by much smaller amounts. Using a "z" test, the decline in correlation coefficients for the aggregated model is the only change that is statistically significant. The relative stability of in-class, take-home, and paper regression models from year 1 to years 1–3—both in terms of correlations coefficients and beta weights—is even more surprising when we recognize that the second and third years of law school are comprised almost entirely of elective courses. In other words, the regression models for each testing method retain roughly the same predictive ability even as the underlying subject matter becomes more diverse.

A second important finding reflected in Table 16 is that, within the national law school sample, the LSAT appears to be a relatively strong predictor of performance on in-class exams and a relatively weak predictor of performance on take-home exams. Further, the LSAT regression weights for papers are not statistically significant at p = .05 during year 1 or years 1–3. Similar to the divergence of correlation coefficients in Phase 1, this pattern among LSAT regression weights is consistent with the hypothesis that speed is a variable that affects performance (i.e., ordinal ranking of students) on both the LSAT and actual law school exams.

The relatively strong predictive power of the LSAT on in-class exams may also explain why the predictive power of the aggregated model declines after year I. Specifically, as the mixture of grading methods migrates away from in-class exams (where the LSAT predicts best) and toward take-home exams and papers during the second and third years of law school (where the LSAT predicts poorly), the aggregated model necessarily becomes a less accurate predictor of law school performance. The aggregated model's

correlation coefficient between predicted GPA_{law} and actual GPA_{law} was 0.404 in year 1 and 0.307 in years 1–3. As evidenced by the proportions shown in Table 17, this decline may be attributed, at least in part, to the changing composition of testing methods.

Table 17: Proportion of Each Testing Method, National Law School Sample

Testing Method	*Year 1*	*Year 2*	*Year 3*	*Years 1–3*
In-Class Exams	74.9%	60.3%	52.1%	61.3%
Take-Home Exams	13.6%	16.0%	15.7%	15.2%
Assigned Papers	11.6%	19.6%	26.6%	20.0%
Clinical Courses	0.0%	2.6%	4.3%	2.5%
Unknown	0.0%	1.5%	1.2%	1.0%

After using the regression models to calculate a predicted GPA for each student on each testing method, we used the disaggregated model to calculate predicted GPA_{law} for each student in the sample. During year 1, the disaggregated model does not predict significantly better than the aggregated model. This result may be attributable in part to the relatively large proportion of in-class exams during year 1. In contrast, the disaggregated model shows a significant improvement over the aggregated model during years 1–3. During this time interval, the correlation coefficient between actual GPA_{law} and predicted GPA_{law} for the disaggregated model is 0.358. This value is significantly higher than the correlation coefficient for the aggregated model (R_A), which is 0.307. As noted earlier, this increased predictive ability occurs because the disaggregated model effectively includes testing method as a variable. Table 18 shows the change in variance (R^2) as we move from UGPA (the best univariate predictor of GPA_{law} in the national law school sample), to the traditional UGPA/LSAT regression model, to the disaggregated model.

Table 18: Change in Variance when Variables are Added to Model, National Law School Sample

	Year 1			*Years 1–3*		
Independent Variables	R	R^2	Δ in R^2	R	R^2	Δ in R^2
UGPA	.323	.104	.104	.262	.069	.069
UGPA & LSAT	.404	.163	.059	.307	.094	.025
UGPA, LSAT & Testing Method	.406	.165	.002	.358	.130	.036

When comparing the change in variance attributable to the LSAT (2.5%) versus the change attributable to testing method (3.6%), it is clear that, at least for the national law school sample, testing method is a significant variable that affects law school performance. Scatter plots for aggregated and disaggregated models for years 1–3 are shown in Appendix 4.

2. Phase 2 Results, Regional Law School.—Table 19 shows the number of students, the standardized regression weights, and correlation coefficients for each testing method within the regional law school sample.

Table 19: Standardized Regression Weights and Correlation Coefficients by Testing Method, Regional Law School Sample

	Year 1				*Years 1–3*			
Testing Method	n	β_{LSAT}	β_{UGPA}	R	n	β_{LSAT}	β_{UGPA}	R
Aggregated Model	584	.473**	.275**	.558**	602	.428**	.328**	.551**
In-Class Exams	580	.467**	.249**	.539**	602	.438**	.306**	.544**
Take-Home Exams	153	.308**	.094	.324**	481	.165**	.181**	.252**
Papers	572	.331**	.311**	.463**	599	.267**	.320**	.426**
1st Yr. Legal Writing	572	.331**	.311**	.463**				
2nd Yr. Legal Writing					566	.095*	.219**	.242**
Papers in Electives					565	.114**	.167**	.207**

* Significant at p = .05, ** Significant at p = .01.

In contrast to the national law school sample, the aggregated model is a fairly robust predictor of law school performance during both year 1 (0.558) and years 1–3 (0.551). The correlations and regression weights are similar to the correlations found by Wightman in a recent, large-scale validity study, thus suggesting that the regional law school is a fairly typical sample.[138] The LSAT and UGPA beta weights for in-class exams, take-home exams, and papers are also fairly stable between the year 1 and years 1–3. Although the changes in regression weights for take-home exams between year 1 and years 1–3 (β_{LSAT}, 0.308 to 0.165; β_{UGPA}, 0.095 to 0.188) are proportionately larger than the changes for other testing methods, this variation is largely attributable to the small number of students who took take-home exams during their first year (153).[139] When applying a "z" test, none of the changes in the beta weights were statistically significant between year 1 and years 1–3. However, when averages in paper courses were disaggregated into first-year

138. *See, e.g.*, WIGHTMAN, *supra* note 51, at 16 tbl.4 (summarizing the correlations coefficient for LSAT, UGPA, and LSAT/UGPA regression model).

139. In the regional law school sample, 153 students took 153 take-home exams during the first year. In contrast, 274 students in the national law school sample took 416 take-home exams during the first year.

legal writing, second-year legal writing,[140] and papers in seminars and elective courses, first-year legal writing was statistically different from second-year legal writing ($p = .01$ for LSAT, $p = .05$ for UGPA) and elective papers ($p = .01$ for LSAT, $p = .01$ for UGPA).

A second important finding reflected in Table 19 is that, within the regional law school sample, the LSAT has the greatest predictive power on in-class exams during year 1 and years 1–3. However, consistent with the hypothesis that test-taking speed is a variable that affects performance on both the LSAT and actual law school exams, the LSAT regression weights consistently declined vis-à-vis testing methods with more generous time limits. Once again, the fact that β_{LSAT} for take-home exams was smaller than β_{LSAT} for papers suggests that something more than a pure time variable may be present. Yet, it is noteworthy that the regression model for papers was dominated by the first-year legal writing courses, which accounted for 44% of all credits earned in courses graded by papers (see Table 20). When the paper courses are further disaggregated, we discover that β_{LSAT} for take-homes during years 1–3 (0.165) is actually higher than the β_{LSAT} for papers written in the second and third years of law school (0.095 and 0.114).

After using the regression models to calculate a predicted GPA for each student on each testing method, the disaggregated model was used to calculate the predicted GPA_{law} for each student in the sample. During year 1 and years 1–3, the disaggregated model failed to predict significantly better than the aggregated model. When comparing actual GPA_{law} and predicted GPA_{law}, the year 1 correlation coefficients for the disaggregated and aggregated models were both 0.558. For years 1–3, the correlation coefficient for the disaggregated model was 0.553 while the coefficient for the aggregated model was 0.551. Scatter plots for both models are contained in Appendix 5.

The relative proportion of each testing method may have been one reason why the disaggregated model failed to emerge as a better predictor. As shown in Table 20, the in-class exam, where the LSAT predicts the best, was the dominant testing method for all years, accounting for 74.6% of all credits earned.

140. At the regional law school, second-year legal writing consisted of two required legal writing and research courses. At the national law school, legal writing was only required during the first year.

Table 20: Proportion of Testing Method, Regional Law School Sample

Testing Method	*% Yr. 1*	*% Yr. 2*	*% Yr. 3*	*% Yrs. 1-3*	*% of all Papers*
In-class exams	80.6%	71.9%	70.6%	74.6%	
Take-home exams	2.70%	8.4%	8.0%	6.2%	
Papers	16.7%	13.3%	10.8%	13.7%	100.00%
1st yr. Legal Writing				6.1%	44.0%
2nd yr. Legal Writing				4.2%	30.3%
Elective Papers				3.4%	25.7%
Clinical Courses	0.0%	6.3%	10.1%	5.2%	
Unknown	0.0%	0.2%	0.5%	.2%	

The relatively high proportion of in-class exams limits the ability of the disaggregated model to provide a more accurate prediction of law school performance. For example, if 100% of all credits are earned in courses with in-class exams, the aggregated model and disaggregated model are, by definition, identical. In the regional law school sample, 74.6% of all courses were graded by in-class exams. In the national law school sample, this figure was 61.3%.[141] Expressing the prevalence of in-class exams as a ratio demonstrates the relative importance of each testing method for predicting law school performance. For example, if 61.3% of all grades are based on in-class exams, as in the national law school sample for years 1–3, the ratio of other testing methods to in-class exams is 0.631 (0.387/0.613). Yet, if the percentage of in-class exams is raised to 74.6%, as in the regional law school sample for years 1–3, the ratio drops to 0.340 (0.254/0.746), thus nearly cutting in half the relative importance of other testing methods in determining law school averages. In this context, it is noteworthy that the proportion of in-class exams for the national law school for year 1 (74.9%) is very similar to the regional law school sample for all years 1–3 (74.6%). Yet, in both cases the disaggregated model failed to produce an improvement over the aggregated model despite substantial differences in the regression equations for each testing method.[142]

141. *See supra* tbl.17.

142. *See supra* tbl.16 (displaying regression weights for the national law school sample) and tbl.19 (displaying regression weights for the regional law school sample).

3. Comparison of National and Regional Samples.—As shown in Table 21, a comparison of the Phase 2 results from the two law school samples provides evidence that the predictive power of the LSAT is closely tied to testing method.

Table 21: Comparison of Regional and National Law Schools by Testing Method

	National				*Regional*			
Testing Method	$ß_{LSAT}$	$ß_{UGPA}$	R	Stronger Predictor	$ß_{LSAT}$	$ß_{UGPA}$	R	Stronger Predictor
Proctored In-Class	.229**	.264**	.373**	UGPA	.438**	.304**	.544**	LSAT
Take-Home Exams	.118*	.215**	.259**	UGPA	.165**	.181**	.252**	UGPA
Papers	.031	.192**	.195**	UGPA	.267**	.320**	.426**	UGPA
1st year Legal Writing	.045	.191**	.194**	UGPA	.331**	.311**	0.463**	LSAT
2nd year Legal Writing	n/a	n/a	n/a	n/a	.095*	.219**	0.242**	UGPA
Seminars, Elective Papers	.031	.167**	.174**	UGPA	.114**	.167**	.207**	UGPA
Clinic Courses	.030	-.078	.000	Neither	.101*	.123**	.158**	UGPA
Aggregated Model	.161**	.240**	.307**	UGPA	.428**	.328**	.551**	LSAT

* Significant at p = .05, ** Significant at p = .01

Within each respective sample, the LSAT has its greatest predictive power on in-class exams, with significantly lower LSAT regression weights on all other testing methods. In contrast, UGPA regression weights appear to be much more stable between testing methods, suggesting that a change in testing method has a smaller effect on the predictive power of UGPA. Further, the LSAT is the weaker predictor (i.e., $\beta_{LSAT} < \beta_{UGPA}$) for all testing methods in the national law school sample and all testing methods except in-class exams and first-year legal writing within the regional law school sample. However, because in-class exams accounted for such a large proportion of all testing methods (74.6%) in the regional sample, the LSAT was also the stronger predictor for the aggregated model (i.e., the traditional LSAT/UGPA regression equation).

One of the most significant differences between the two samples is the large variability of how well the LSAT predicts performance on writing

assignments. Within the national law school sample, the LSAT was uncorrelated with performance on papers. In contrast, within the regional law school sample, the LSAT was a fairly useful predictor of performance in first-year legal writing courses. As mentioned earlier, this outcome may be attributable to range restrictions in the national law school sample, along with the grading criteria that may be unique to first-year legal writing.[143] Yet, it is important to note that when paper grades are disaggregated, the LSAT and UGPA regression weights for seminars and elective papers are very comparable between the two law schools.

4. Summary of Phase 2 Results.—Phase 2 produced different results for each sample group. For the national law school, the disaggregated model was a significantly better model of law school performance than the aggregated model. For the regional law school, the disaggregated model offered virtually no improvement over the aggregated model.

Notwithstanding these differences, an inspection of the regression weights and correlation coefficients suggests that testing method is a variable that affects the ordinal ranking of law school performance. Specifically, in both sample groups, the predictive power of the LSAT is closely tied to in-class exams. On other testing methods, the LSAT regression weights drop off significantly. Although UGPA produced fairly modest correlations with law school performance in both the national and regional law school samples, the regression weights for UGPA were generally much more stable than the regression weights for LSAT. On testing methods other than in-class exams, UGPA also tended to be the stronger predictor. Finally, with the exception of papers and first-year legal writing in the regional law school sample, the regression equations for each testing method were relatively stable between year 1 and years 1–3. This result occurred despite the fact that students take the same required courses in year 1 and a diverse range of electives in years 2–3.

The findings of Phase 2 also suggest that the proportional use of each testing method affects the relative importance of the LSAT and UGPA as predictors of law school performance. Thus, the reason the disaggregated model predicted better in the national law school sample is that in-class exams became a less common grading method (year 1, 74.9%; year 2, 60.3%; year 3, 52.1%).[144] In contrast, the proportion of in-class exams in the regional law school sample remained consistently high during all three years of law school (year 1, 80.6%; year 2, 71.9%; year 3, 70.6%).[145] In an environment in which testing method appears to be a variable that affects law

143. *See supra* notes 135–36 and accompanying text.

144. *See supra* tbl.17.

145. *See supra* tbl.20.

school performance, it is significant that the LSAT has its greatest predictive power on the most common testing method.

C. *Additional Findings Related to Test-Taking Speed*

The results of Phases 1 and 2 generally support the hypothesis that test-taking speed is a variable that affects the performance (i.e., ordering) of students on both the LSAT and actual law school exams. This subpart presents the results of three subsidiary analyses of the data set that are designed to provide additional evidence to either corroborate or rebut the presence of a test-taking speed variable. A fourth analysis explores possible differences in test-taking speed among ethnic subgroups.

1. Differential Between In-Class Exams and Other Testing Methods.— For each student in the national and regional law school sample, we assigned an ordinal ranking according to their GPA on in-class exams, take-home exams, papers, and take-home exams and papers combined. We then created three new variables that reflected the differential (D) between the ranking on in-class exams and the ranking in one of the other categories:[146]

$$D_1 = Rank_{take\text{-}home} - Rank_{in\text{-}class}$$

$$D_2 = Rank_{papers} - Rank_{in\text{-}class}$$

$$D_3 = Rank_{take\text{-}home\ \&\ papers} - Rank_{in\text{-}class}$$

It was hypothesized that students with low test-taking speed would tend to have large negative differentials due to (1) a lower ranking on in-class exams, where test-taking speed may be important, and (2) a higher ranking on take-home exams and papers, where the effect of test-taking speed is presumably diminished or negligible. For example, if a student's ranking on take-home exams was 10 and his ranking on in-class exams was 100, the ranking differential (D_1) for that student would equal 90. Further, if test-taking speed is a variable that affects performance on both the LSAT and law school grades, it was expected that this differential would be positively correlated with the LSAT. The results are presented in Table 22.

146. The ordinal ranking for in-class exams corresponded to the number of students who had taken the other testing method. For example, when calculating "Take-Home Rank minus In-Class Rank" in the regional law school sample, ranking for in-class exams was limited to the 482 students who had actually taken a take-home exam. This technique guaranteed that the mean differential for each comparison would be zero. Thus, net gains in rank perfectly counterbalanced net declines.

Table 22: Correlation between Differential Ranking Variable (D) and LSAT, UGPA

	National Law School			*Regional Law School*		
Differential (D)	n	LSAT	UGPA	n	LSAT	UGPA
Take-Home Rank Minus In-class Rank	377	.112*	.064	482	.245**	.113*
Paper Rank Minus In-class Rank	379	.199**	.108*	601	.174**	-.008
Take-Home & Paper Rank Minus In-class Rank	379	.171**	.095	603	.174**	-.007

* Significant at p = .05, ** Significant at p = .01

Although the correlations between LSAT and the differential ranking variables (D) are relatively modest, they are statistically significant and comparable between the two samples. Furthermore, in both samples, the differential between take-home exams and in-class exams was strongly correlated with the differential between papers and in-class exams (national, 0.412; regional, 0.366). The most plausible explanation for these results is that the LSAT is a measurement (in proportions that vary with each student) of both reasoning ability and test-taking speed.

2. Relationship Between LSAT and Page Length of Exam.—The record-keeping at the national law school permitted visual inspection of virtually all of the exams in the sample. Thus, the variable of exam page length was added to the data set.[147] We hypothesized that students with low test-taking speed would tend to do worse on exams with a greater number of total pages. Of the 11,312 individual grades in the sample, 6,236 (55.13%) were based on in-class exams using a pure essay format. Of this group, 2,767 (44.4%) were four pages or less in length, and 3,469 (55.6%) were five pages or more in length. The division between four and five pages offered the proportion closest to 50%. The LSAT and UGPA coefficients for these categories are presented in Table 23.

147. Most tests were single-spaced, 12 point font. When an exam was doubled-spaced, the length was divided in half and one page was added. Thus, an eight-page, double-spaced exam was coded as five pages (8/2 +1).

Table 23: LSAT & UGPA Correlation Coefficients
for Long Versus Short In-Class Essay Exams

Category	*n*	*LSAT*	*UGPA*
Short In-Class Essay (≤ 4 pages)	2,767	.190**	.287**
Long In-Class Essay (≥ 5 pages)	3,469	.270**	.271**
Other Testing Methods in Sample	5,076	.145*	.217**
All Testing Methods in Sample	11,312	.194**	.262**

* Significant at p = .05, ** Significant at p = .01

Using a "z" test, the LSAT correlation coefficient for short exams (.190) is statistically different from long exams (.270) at p = .01. This finding suggests that test-taking speed is a variable common to both the LSAT and longer (in terms of number of pages) in-class essay exams.

3. Evidence Based on Age.—As noted in the literature review, test-taking speed typically declines with age after early adulthood.[148] Because only a small proportion of students in the national and regional law school sample entered law school when they were age 28 or older (national, 9.7%; regional, 16.9%), the influence of age-related differences in test-taking speed is likely to be small. Nonetheless, the test-taking speed hypothesis draws limited additional support when the variable of age is considered.

In the national law school sample, entering age was uncorrelated with LSAT score and negatively correlated with UGPA (-0.356). Thus, based on the importance of UGPA in the regression equations, if age is not a variable affecting law school performance, age should be either uncorrelated or negatively correlated with GPA on all testing methods. Nevertheless, entering age was positively correlated at statistically significant levels with both GPA_{papers} (0.106, p = 0.05) and $GPA_{take\text{-}home\ \&\ papers}$ (0.110, p = 0.05). Moreover, when age is included in the LSAT and UGPA regression models for predicting performance on assigned papers, the correlation coefficient increased from 0.195 to 0.271. Similarly, when age was added to a regression model for grade averages on take-home exams and assigned papers combined, the model's correlation coefficient increased from 0.229 to 0.305.

In the regional law school sample, entering age was uncorrelated with LSAT and negatively correlated with UGPA (-0.115). In contrast to the national sample, entering age was not correlated with GPA_{papers} or $GPA_{take\text{-}}$

148. *See supra* note 111 and accompanying text.

$_{\text{home \& papers}}$ at statistically significant levels. However, in a forward stepwise regression that excluded all variables that failed to improve the model at statistically significant levels ($p = .10$), age was excluded from the model for $GPA_{\text{in-class}}$ but included for GPA_{papers} and $GPA_{\text{take-home \& papers}}$. With age added as a variable, the correlation coefficient (R) for GPA_{papers} increased from 0.426 to 0.434; the correlation for $GPA_{\text{take-home \& papers}}$ increased from 0.431 to 0.442.

Although these results may be attributable to age-related differences in test-taking speed, an alternative theory is that writing ability is a distinctive skill that improves with experience over time. Therefore, older students may have a slight advantage on certain testing methods, such as papers and take-home exams, which may place a greater premium on a clear, well-written analysis.

4. Evidence Based on Ethnicity.—As noted in the literature review, there is some evidence to suggest that ethnic subgroups, particularly Hispanics and African Americans, may be disproportionately affected by a speededness component on the LSAT.[149] The results of this study yielded inconclusive evidence that the testing method, defined by time allowed, has a differential impact on minority students.

The performance of students by ethnic subgroup was analyzed using the differential ranking variables (D) discussed above.[150] If minority students are more likely to be adversely affected by time pressure, we would expect higher ordinal rankings based on $GPA_{\text{take-home}}$, GPA_{papers} or $GPA_{\text{take-home \& papers}}$ and lower ordinal rankings based on $GPA_{\text{in-class}}$. Thus, in terms of the differential ranking variable, which subtracts $Rank_{\text{in-class}}$ from $Rank_{\text{take-home}}$, $Rank_{\text{papers}}$, or $Rank_{\text{take-home \& papers}}$, if students within an ethnic subgroup tend to be adversely affected by stringent time limits, then the mean differential for that subgroup should be negative. As shown in Table 24, in both samples, white students tend to have positive differentials (i.e., ordinal ranking is higher on in-class exams than on other testing methods), and minority students tend to have negative differentials (i.e., ordinal ranking is lower on in-class exams than other testing methods). In order to visualize the distributions for each subgroup, Appendix 6 contains box plots for the change in ranking between take-home and in-class exams for both the national and regional law school samples.

149. *See supra* notes 98–99 and accompanying text.

150. *See supra* section IV(C)(1).

Table 24: Mean Differential of Ordinal Ranking by School and Ethnicity

			Mean Differential of Ordinal Ranking		
School	*Ethnicity*	*n*[151]	*Take-Home Minus In-Class*	*Paper Minus In-Class*	*Take-Home & Paper Minus In-Class*
National	*White*	263	6.89	4.80	4.59
	Asian	33	-23.51	4.85	-5.94
	Black	13	-66.62	-14.69	-31.69
	Hispanic	18	18.78	-12.22	14.28
	Undeclared	50	-10.18	-23.80	-17.38
Regional	*White*	388	6.50	5.46	4.39
	Asian	23	4.09	-47.73	-29.00
	Black	25	-17.16	-39.61	-14.00
	Hispanic	16	-86.44	-25.58	-32.63
	Undeclared	30	-26.80	14.77	-1.51

The methodology used to generate the mean differential variables displayed in Table 24 has some advantages that facilitate comparisons between subgroups. For example, as a matter of pure math, the sum of the mean differential variables must equal zero.[152] Therefore, when the 263 white students at the national law school averaged a 6.89 gain in relative ranking on in-class exams (a speeded environment) versus take-home exams (an unspeeded environment), an aggregate loss of ranking units of 1,812

151. The number of observations in this column only reflects students who took take-home exams. In the national law school sample, only two students did not take a take-home exam during their three years of law school. In contrast, only 482 out of 604 students in the regional law school sample took take-home exams. The numher of students who wrote papers essentially overlapped with the number of students in each sample. All students in the national sample and all but two students in the regional sample wrote at least one paper but usually wrote several papers.

152. *See supra* note 146 (explaining that the mean differential for take-home exams minus in-class exams was limited to students who actually took both methods, thus guaranteeing that both the sum and the average of the differentials would always equal zero).

(6.89 x 263) has to be distributed over the remaining students, who are either minority or undeclared status. In other words, comparisons of these variables will always result in a zero-sum game; for one group to gain, another must experience a loss.

Despite a clear trend toward positive differentials for white students and negative differentials for minority students,[153] the small number of minority students in both sample groups precludes any definitive finding that minority students perform disproportionately worse on in-class exams. However, even with the small sample size, some of the differentials between white students and minorities were statistically significant. For example, the mean differential between take-home and in-class exams for black students in the national law school sample (-66.62) was found to be statistically different from the mean differential for white students (6.87) at $p = .05$. The mean differential between take-home and in-class exams for Hispanic students in the regional law school sample (-86.44) was also found to be statistically different from the same figure for white students (6.50) at $p = .05$.

The differential between take-home and in-class exams is arguably the best category to examine subgroup differences because blind grading was applied to both testing methods in both samples (see Appendix 6 for box plots). Yet, it should be emphasized that the trends for Hispanic and black students were different between the two sample groups. In the national law school sample, the mean differential between take-home and in-class exams for Hispanic students was positive (18.77). Moreover, in the regional law school sample, the mean differential between take-home and in-class exams for black students (-17.16) was not statistically significant. Thus, at a minimum, a larger sample size needs to be obtained in order to adequately

153. Professor Guinier has reported anecdotal evidence that females tend to perform better on take-home exams and research papers than traditional law school exams. *See* Lani Guinier, *Lessons and Challenges of Becoming Gentlemen*, 24 N.Y.U. REV. L. & SOC. CHANGE 1, 7–8 (1998) (reporting the observation of colleagues that "many women perform better on take-home exams and research assiguments that give them ample opportunity to think and reflect"). The results of this study failed to provide any significant support for this hypothesis. In the national law school sample, females actually had better ordinal rankings on in-class exams. These results are summarized in the following table (n is limited to students who actually took take-home exams, see *supra* note 151).

School	*Gender*	*n*	*Take-Home minus In-Class*	*Paper minus In-Class*	*Take-Home/Paper minus In-Class*
National	*Male*	232	-2.42	-2.72	-3.97
	Female	145	3.87	4.34	6.34
Regional	*Male*	242	-0.88	12.18	8.49
	Female	240	0.88	-12.18	-8.52

explore the possibility of a testing method differential among ethnic subgroups.[154]

V. Implications of this Study

The primary contribution of this Article is empirical. The evidence presented in Part III provides a broad array of evidence that test-taking speed is a variable that affects student performance on the LSAT and actual law school testing methods. Moreover, if the results from the national and regional law schools in this study are fairly representative of other law schools, it follows that the legal academy's heavy and under-theorized reliance on time-pressured exams exerts a significant (and heretofore unrecognized) effect on the relative importance of the LSAT as an admissions criterion.

The objective of Part V is to provide a preliminary response to at least three areas of uncertainty created by this study. Subpart A describes the precise mechanism of how a speed variable that affects both the predictor (LSAT) and the criterion measure (actual law school performance) can increase the predictive validity of the LSAT. Subpart B then focuses on the relationship between "speed" and the legal profession. Notwithstanding my skepticism that time-pressured law school exams provide an accurate analogue to the practice of law,[155] it is certainly *not* my claim that talented lawyers who work slowly are on par with talented lawyers who work quickly—efficiency is certainly relevant. Nor do I dispute that lawyers who possess a "quick wit" will often serve their clients well.[156] Rather, subpart B

---

154. This study merely considers the possibility that a test-taking speed differential might exist between ethnic subgroups. A separate but important question is *why* a differential might be present. The following passage from Professor Hill suggests at least one plausible explanation:

> People who come out of a reading tradition have a greater affinity for taking examinations like the LSAT than those who do not. People who come out of an oral tradition [such as African Americans] are therefore at a distinct disadvantage. This does not mean, however, that people who come out of a reading tradition are smarter or, are more creative than those from an oral tradition. Nor does it mean that those from an oral tradition are culturally inhibited from mastering the art of reading. It merely reflects group values, and the process of group socialization.

Donald K. Hill, *Law School, Legal Education, and the Black Law Student*, 12 T. MARSHALL L. REV. 457, 489 (1987). *See also infra* subpart V(B) (summarizing evidence from regional law schools that scores on oral portions of appellate advocacy classes were *uncorrelated* with LSAT, UGPA, or overall law school GPA, whereas the correlation coefficient for brief writing and overall GPA was .40).

155. *See supra* notes 23 and 34 and accompanying text.

156. *Cf.* Kisman, *supra* note 24, at 454. Kisman notes that

> [M]any professors appear to believe that all good lawyers must have the poise and quick wit to deal with continuous surprises because good appellate advocates are believed to possess these qualities. Professors believe in the benefits of these qualities, of course, because they showed such poise and quick wit on their own Blue Book exams.

Id.

discusses the conceptual and empirical difficulties of asserting that the LSAT or time-pressured law school exams provide a reliable proxy for these virtues. Finally, because the data presented in Part III suggests the need for a new framework for predicting and interpreting law school performance that goes beyond the LSAT and UGPA, subpart C outlines a slightly broader theory that fits the data contained in the regional and national law school samples.

A. Predictive Validity on the LSAT: The Relationship Between Speed and Range Restriction

As a threshold matter, it should be noted that the presence of a speededness component on a power test does not necessarily render the results of that test unreliable. In general, the results of a speeded power test will be very strongly correlated with the same test given under unspeeded conditions.[157] Thus, it could certainly be argued that any "speededness confound" on the LSAT would have only a *de minimis* effect on the test's predictive power. However, if speed is a significant variable that affects both the predictor (LSAT) and the criterion measure (actual law school grades), it is possible that this speededness component may actually *increase*, rather than undermine, a test's predictive validity.[158] Further, this inflationary effect is likely to be the most pronounced in a range restricted environment—a condition which is present, in varying degrees, in virtually all LSAT validity studies.[159] As a result, differences in test-taking speed rather than reasoning ability may account for why the LSAT typically emerges as a stronger predictor of law school performance than UGPA.[160]

---

157. *See, e.g.*, KLINE, *supra* note 17, at 72 (noting that "there is a high correlation between the same [power] test, timed and untimed"); Simon G. Draycott & Paul Kline, *Speed and Ability: A Research Note*, 17 PERSONALITY & INDIVIDUAL DIFFERENCES 763, 767 (1994) (noting that variations on timed intelligence tests appear to be explained primarily by differences in ability rather than differences in speed); Baxter, *supra* note 81, at 285 (noting that "under most conditions," the correlations between timed versus untimed intelligence tests are typically "fairly high, often in the .90s").

158. *See, e.g.*, *supra* note 81.

159. In an LSAT validity study, correlation coefficients for LSAT and UGPA are generated for each individual law school. However, the majority of students at virtually all law schools fall within a fairly narrow range of LSAT scores. For example, in the 2003 survey in *Best Graduate Schools*, U.S. NEWS & WORLD REP., 2003, the difference between the 75th percentile LSAT score and the 25th percentile LSAT score broke down as follows: three point difference at 4 schools; four points at 21 schools; five points at 47 schools; six points at 52 schools; seven points at 30 schools; eight points at 15 schools; nine points at 4 schools; ten points at 1 school. Yet, notwithstanding the range of LSAT scores at any particular law school, at the top end of the rankings, virtually the entire student population is drawn from students who scored in the 90th or 95th percentile overall (i.e., a very homogeneous group in terms of reasoning ability). *Special Issue: Best Graduate Schools, Schools of Law*, U.S. NEWS & WORLD REPORT, Apr. 15, 2002, at 64–68.

160. For sources documenting the stronger predictive power of the LSAT vis-à-vis UGPA, see *supra* notes 51–53 and accompanying text.

It is important to understand how this inflationary mechanism operates. Specifically, the LSAT appears to contain two pieces of information—reasoning ability and test-taking speed—which vary in relative proportion for each individual test-taker.[161] As just noted, if the LSAT is a speeded exam, it may still provide a fairly accurate method for arraying candidates by reasoning ability, albeit less accurately than the same test given under less speeded conditions. This result occurs because candidates with fast-taking speed will likely achieve *marginally* higher LSAT scores, whereas candidates with a low test-taking speed will likely achieve *marginally* lower LSAT scores.[162] Because the majority of students at most U.S. law schools fall within a fairly narrow range of LSAT scores, these student populations will tend to be relatively *homogeneous* in terms of "acquired verbal reasoning and reading skills," which is the construct the LSAT is designed to measure. However, because test-taking speed is uncorrelated with reasoning ability,[163] most law school populations will also tend to be relatively *heterogeneous* in terms of test-taking speed.[164]

A simplified example illustrates how the inclusion of a speed component on both the LSAT and traditional law school exams inevitably increases the predictive validity of the LSAT. Assume that at the national law school, several candidates who have high reasoning ability and fast-test-taking speed receive a 171 on the LSAT (75th percentile within the sample, ~99th percentile overall). In contrast, several candidates who have the same level of reasoning ability but slow test-taking speed receive a 165 (25th percentile within the sample, ~94th percentile overall).[165] If law school testing methods were unspeeded (e.g., take-home exams and papers), then, all other factors being equal, both groups of candidates 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

161. The stylized example of Anderson and Benton in subpart II(A) is designed to illustrate this relationship. It should be noted, however, that the LSAT is designed to be a univariate (i.e., single variable) test that measures a specific construct ("acquired verbal reasoning and reading skills"). The introduction of a speededness component therefore makes the LSAT a composite of at least two discrete, separable mental abilities: reasoning skills and speed. As a result, a candidate who is "fast and fairly accurate" may get the same score as another candidate who is "slow and very accurate." *Cf.* KLINE, *supra* note 17, at 35 ("[I]t is essential that all [psychometric] tests are univariate . . . . A test which measures more than one variable cannot be accurately interpreted and identical scores may not be psychologically the same.").

162. In other words, unless the speed component overwhelms the underlying construct that the power test is designed to measure, the change in percentile ranking is unlikely to be dramatic. For this reason, speeded power tests are strongly correlated with the same test given under unspeeded conditions. *See supra* note 157.

163. *See supra* note 18.

164. For example, a candidate with high reasoning ability and fast-test-taking speed may receive a 171 on the LSAT, whereas another candidate with the same level of reasoning ability but slow test-taking speed may receive a 165. Yet, both candidates score high enough to get into the same elite law school.

165. *See supra* tbl.4 (showing LSAT scores for students who entered as 1Ls).

testing method,[166] then the ordering of test-takers will tend to track the ordering of the LSAT, thus driving up the LSAT's predictive validity. Yet, in this example, the higher correlation coefficient will be attributable to differences in test-taking speed rather than differences in reasoning ability.

In terms of the present study, it is noteworthy that the LSAT consistently best predicts performance on in-class exams, with significantly lower correlation coefficients on take-home exams and papers.[167] It may therefore be the case that high LSAT correlation coefficients, which usually hover in the 0.4 to 0.45 range in most LSAT validity studies,[168] may not be possible without time-pressured, in-class exams as the dominant testing method. This issue is extremely important because high LSAT correlation coefficients vis-à-vis lower UGPA correlation coefficients (which usually hover in the 0.2 to 0.3 range)[169] mean that the LSAT will usually be given more weight during the admission process. While it may be normatively unappealing to disadvantage slow but talented test-takers, the linkage between speed and the relative weighting of the LSAT also has a disparate impact on minority candidates. Specifically, the performance gap between white and minority students tends to be larger on the LSAT than on UGPA.[170] Thus, heavy reliance on in-class exams will tend to produce LSAT/UGPA admission indexes that will admit fewer minority students through the regular admissions process.[171] Of course, this relationship begs

---

166. In the national law school sample, the in-class exam was the principal grading method for 74.9% of first year grades and 61.3% of the grades for years 1–3. *See supra* tbl.17. The median time limit was 3 hours and the average time limit was 2 hours, 52 minutes. *See supra* tbl.8.

167. *See supra* tbl.9 (showing the national law school sample), tbl.10 (showing the regional law school sample).

168. *See* sources cited *supra* notes 51–53.

169. *Id.*

170. *See* sources cited *supra* note 32.

171. When I use the term "regular admission process," I am referring to a process that relies *in part* on an "admissions index." *See* LSAC REFERENCE MANUAL, *supra* note 16, at 16. This index is the product of a two-variable linear regression that establishes the relationship between LSAT and UGPA (the predictor variables) and first-year law school grades (the criterion variable). *Id.* at 24. The Office of Student Academic Services at Oklahoma State University provides its undergraduates with the following useful description of how the index is typically used:

> Law School Data Assembly Serviee [run by the LSAC] calculates an admission index using the undergraduate gpa and the LSAT score statistically weighted. This calculation is based on each school's instructions for the weighting that best predicts first year success at that school. Many schools may use this index to establish some broad admissions guides.

Admissions Requirements and Considerations, *available at* http://www.cas.okstate.edu/services/campus/law/reqs.html (last visited Jan. 10, 2004). Similarly, the Pre-Law Advising Office at the University of Massachusetts offers its students the following description of how the index is used in the sorting process that occurs at most U.S. law schools:

> Most law schools use an index to rank applicants .... Based on the index, the applicant pool is divided into three categories: presumptive admits, presumptive denies, and all the rest. The presumptive admits will generally be accepted barring any negative information in the file. Likewise, the presumptive denies will generally be rejected unless there is some compelling information in the file. The majority of

the question of whether there is sound educational, theoretical, or empirical basis for emphasizing speed on law school exams. Subpart B focuses directly on this issue.

B. "Speed" and the Legal Profession

Assuming that the above framework accurately describes the linkage between time-pressured exams and the relative weight assigned to the LSAT, the legal academy may need to consider which testing method (in-class exams, take-home exams, papers, or some other category) is the best measure of legal ability and accomplishment. As mentioned earlier, it could be argued that papers and take-home exams are a closer analogue to the practice of law, in terms of both time pressure and the creation of final work product that might be relied upon by a client, another lawyer, or a court.[172] Alternatively, it is hard to deny the intuition that some raw quantity of "speed" or "mental quickness" is a characteristic that might be valuable in a variety of legal contexts.[173]

---

applicants in the pool will fall somewhere in the middle. These candidates are compared to each other using more subjective criteria.

Law School Decision Process, *available at* http://www.umass.edu/uaasc/prelaw/decisions.htm (last visited Jan. 10, 2004). The importance of numerical criteria can also be observed in the LSAC's online database, which estimates the likelihood of admission to specific law schools based on UGPA and LSAT credentials. *See* http://officialguide.lsac.org/search/cgi-bin/lsatgpa.asp (last visited Jan. 10, 2004).

Assuming that admissions indexes are important in law school admissions, the following claim should not be controversial: If the LSAT becomes a less important predictor of law school success, minority students will tend to fare better under the current system of admissions. Because the performance gap between white and minority students is larger on the LSAT than UGPA, see *supra* note 34, a migration away from less-time pressured law school testing methods will, on average, have the effect of improving the admission indexes of minority students. In other words, more minority students will move into the presumptive admit and middle (or discretionary admit) categories. Although it is doubtful that a change in law school testing methods would, by itself, be sufficient to achieve a "critical mass" of underrepresented minorities, Grutter v. Bollinger, 123 S. Ct. 2325, 2342–44 (2003) (discussing the relevance of "numbers" to achieve the educational benefits of a diverse student body), a reduction in the underlying numerical gap would make this process less contentious.

172. *See supra* note 34 and accompanying text.

173. In 1926, the eminent psychometrician Edward L. Thorndike opined, "If speed deserves any weight in determining the measures of intellect it is by virtue of the principle that, 'Other things being equal, the more quickly a person produces the correct response, the greater is his intelligence.'" CARROLL, *supra* note 18, at 440 (citing EDWARD L. THORNDIKE ET AL., THE MEASUREMENT OF INTELLIGENCE 24 (1926)). However, in contrast to standardized tests, responses to legal problems do not neatly divide into "correct" and "incorrect" answers. Rather, there is often a continuum between good, better, and brilliant. The issue raised by this study is whether, given more time, some students would give a *more* "correct" answer than students who run quickly out of the gate. When it comes to brilliance, the law is sometimes willing to be patient. An eminent psychometrician of the modern era, John B. Carroll, later clarified that Thorndike's observation reflected a "societal judgment" on intelligence rather than a statement of scientific fact; empirical evidence amassed after Thorndike's comment demonstrated that "speed" and "level" (i.e., reasoning ability) were, in fact, distinct and separate abilities. *Id.* at 508–09.

Time is certainly relevant in the legal profession. Lawyers bill by the hour. They are also occasionally pressed by clients to provide immediate legal advice over the phone without the benefit of any research or reflection. An objection to an evidentiary issue cannot be the subject of an appeal unless it has been *timely* raised before the trial court.[174] Similarly, appellate judges pride themselves on raising novel and unexpected issues during oral argument.[175]

The difficult analytical question considered here, however, is whether the time facility required on the LSAT or law school exams is an accurate metric for these widely divergent concepts of efficiency and speed.[176] To further complicate the analysis, the examples just cited can be counterbalanced by a wide range of legal tasks that are performed over a period of days, weeks, or months.[177] Thus, an important subsidiary issue is whether it is appropriate to use time-pressured exams as the *dominant* testing mode. For the purposes of this section, I will divide "speed" into two analytical categories: (1) efficiency in generating a quality written work product, and (2) intellectual agility or quickness in a verbal exchange, such as an oral argument.

The following hypothetical serves to sharpen the issue of speed in the context of legal writing. Imagine that a partner at a law firm calls a junior associate into his office at 8:00 am. He asks her to write a memo on an important legal issue. The partner clearly wants the memo quickly, but he

---

174. *See* FED. R. EVID. 103(a).

175. As a former judicial clerk for a U.S. Circuit Court of Appeals, it was my experience that oral argument rarely changed the impression of a case that emerged after reading the briefs. In contrast, weeks after oral argument, votes would occasionally change in response to additional research done within chambers; these situations certainly deepened my faith in the fairness of the judicial process. Harmonizing a case with the existing body of law is a task that requires a great deal of precision. Despite its aesthetic value, the legal repartee of oral argument does not always advance this goal. To my mind, the value of oral argument is to clarify points that were unclear from the briefs. That said, I readily concede that other law clerks might have had a different experienee.

176. *Cf.* Wesman, *supra* note 66, at 267. Wesman questions the usefulness of speeded tests in the employment context:

> One reason for choosing the speeded test is . . . that the firm wants men who can think quickly and can make rapid decisions. The implicit assumption is that speed as represented in the test and speed as represented on the job are one and the same. It is a dubious assumption at best.

Id.

177. For example, a polished appellate brief or motion for summary judgment is often written over a period of weeks or months. Likewise, a client who has a complex tax problem or a high-stakes, corporate transaction is unlikely to want a speedy work product at the expense of accuracy. Thus, hundreds of billable hours are expended late into the night in an effort to eliminate potential errors. The prosecution of recent corporate scandals highlights the fact that some important legal issues are just not amenable to a speedy resolution. *See, e.g.*, *Former Enron Executive's Trial Delayed*, N.Y. TIMES, Feb. 11, 2003, at C4 (reporting on the granting of a 90-day delay in prosecution of former Enron CFO Andrew Fastow, "because of the huge amount of paperwork that has been generated . . . . 'There's no good way to make it simple.'" (quoting an assistant U.S. attorney working on the case)).

fails to state a specific deadline. The associate then retreats to the library to research and write the memo. At 11:30 am, the associate is summoned to the partner's office and asked for a draft. The associate hands over 10 pages that resemble a typical law school Blue Book exam. The analysis is starting to develop but the prose is awkward and the organization is poor. The partner sends the associate back to the library with the instructions, "make it better, not longer." At 5:00 pm, the associate is called back to the partner's office and asked for a revised draft. Similar to an eight-hour take-home exam, the organization is easier to follow and the legal analysis is crisp and more compact.[178] The partner has renewed confidence in the associate's ability. However, there are still several loose ends, and the matter is now going to litigation. The partner instructs the associate to turn the memo into a legal brief, nailing down all the finer points. A week later, the associate turns in a polished jewel, similar to an "A" paper in law school.

This hypothetical illustrates the fact that the memo delivered at 11:30 am is a mere snapshot of a work in progress. An analysis hurriedly prepared in three hours may or may not be indicative of the work product that might arrive at the end of the day or the end of the week.[179] The empirical evidence from this study strongly suggests that the ordering of test takers can change dramatically when a different testing method is used.[180] The national law school sample provides a vivid example of this outcome. As noted earlier, a GPA for each testing method (in-class exams, take-home exams, and papers) was calculated for each student. Of the 379 students in the sample, the student with the highest $GPA_{in\text{-}class}$ also had the highest LSAT score in the sample. In contrast, the student with the highest GPA_{papcr} had an LSAT score in the bottom 2% of the sample. Surprisingly, this student also had a below average $GPA_{in\text{-}class}$ and an average $GPA_{take\text{-}home}$. Yet, in the context of paper

178. *See supra* note 126 (noting that faculty members at national law schools give take-home exams with a word limit precisely because they produce answers that are better written and organized than are in-class exams).

179. *Cf.* K. Robert Bridges, *Test—Completion Speed: Its Relationship to Performance on Three Course-Based Objective Examinations*, 45 EDUC. & PSYCHOL. MEASUREMENT 29, 33–34 (1985) (comparing results and time of performance on three untimed objective tests and finding that "[s]tudent performance could not be predicted by the relative order of test completion or by the time required to complete a test").

180. The findings of this study rebut the assumption, held by some in the academy, that testing method has no effect on the ordering (i.e., relative grade) of students. For example, Professor Weaver has made the following observation, which now appears to be erroneous:

> Does the take-home exam produce different outcomes? In other words, do some students do better on the take-home exam than they would have done on a traditional final exam? My sense is no. Students who do well (or poorly) on a traditional final exam do similarly on a take-home exam. The reasons are obvious. Students do better on final exams because they have better reasoning abilities and better writing skills. Students who do poorly generally lack these attributes. These differences reveal themselves in either type of testing mechanism.

Russell L. Weaver, *Teaching (and Testing) Administrative Law*, 38 BRANDEIS L.J. 273, 287 (2000).

writing (and this student wrote many papers),[181] this individual was clearly exceptional.[182] The point of this anecdote is that overreliance on the LSAT or performance on in-class exams can produce some egregious errors.[183] Assuming that a lawyer must produce a work product that can be relied upon by a client, another lawyer, or a court, time-pressured law school exams appear to be an unrealistic metric for efficient writing ability.

Turning now to the context of verbal facility, the relevant question is whether "speed" or "quickness" on the LSAT or in-class exams is a useful proxy for the skills necessary in oral advocacy. In other words, the question is whether a student's ability to bubble in the correct answer on the LSAT within 1.25 minutes, or write a cogent in-class exam under intense time-pressure, is also a valid predictor of a student's eventual performance as an oral advocate.[184]

The regional law school sample provided some limited empirical evidence on this issue. During the sample period, the regional law school enrolled forty-one students in an advanced class on appellate advocacy. Each section was taught by the same instructor. The final grade for the class was based 50% on a written brief, and 50% on a series of oral arguments. Although the sample size was quite small, the results were rather surprising. For example, the grades on brief writing were poorly correlated with oral argument grades (0.121), which suggests that oral skills may be quite distinct from writing ability. Moreover, neither the LSAT nor UGPA had a statistically significant correlation with either the grades for brief writing or the grades for oral argument.[185] Furthermore, two of the three students with the

---

181. In fact, this student wrote more papers than anyone else in the sample. This suggests that the student actively sought out paper writing, presumably because this was an area of comparative advantage.

182. Moreover, this student was not necessarily an outlier. Overall, many students with relatively low LSAT scores had very high GPA_{papers}. Not surprisingly, LSAT scores had zero correlation with paper grades for the sample as a whole. *See supra* tbl.9.

183. In contrast, this anecdote is not an indictment of the admissions process. The student had unexceptional entering credentials (the individual's UGPA was also significantly below the median). Apparently, the admissions officer relied on other indicia of talent or ability. In other words, the "whole person" approach seemed to work.

184. It should be noted that the Law School Admission Council emphatically cautions against the use of the LSAT for any purpose other than predicting first-year grades. *See, e.g.*, LSAC REFERENCE MANUAL, *supra* note 16, at 15. The LSAC Research Manual notes that:

> The LSAT is designed to serve only the admissions process and has not been validated for any other purpose. LSAT performance is subject to misunderstanding and misuse in other contexts, as in making an employment decision about an individual who has completed most or all of his or her law school work.

Id. Notwithstanding this advice, the LSAT is sometimes used as a hiring criterion by federal judges and law firms. *See* William C. Kidder, *Portia Denied: Unmasking Gender Bias on the LSAT and Its Relationship to Racial Diversity in Legal Education*, 12 YALE J.L. & FEMINISM 1, 22–24 & nn.103–06 (2000) (collecting evidence on unintended uses of the LSAT).

185. Although none of the correlation coefficients were statistically significant, which is not surprising based on the small sample size, the highest correlation existed between UGPA and brief

highest oral argument scores had the second and third *lowest* LSAT scores among the forty-one students—nearly twenty points lower than the top LSAT score. Although it might be argued that these results were the product of a highly subjective grading process, it is noteworthy that the brief writing scores had a fairly strong and statistically significant correlation to the students' final law school GPAs (0.406, p = .01). In contrast, the correlation between the students' oral scores and final GPAs was positive (0.229), but below the level of statistical significance for the sample size (n = 41).

In summary, the analysis of the appellate advocacy sample suggests that the "speed" or mental quickness that might aid a test taker on the LSAT or time-pressured law school exams may, in fact, be completely *unrelated* to oral skills that are prized within the courtroom. Thus, it would appear that the academy's current emphasis on time-pressured testing methods (both for admission and for grading) may lack both a theoretical and an empirical justification.

In light of the Supreme Court's ruling in *Grutter*, this is hardly a matter of passing interest. The results of this study suggest that the predictive validity of the LSAT is inextricably tied to the legal academy's heavy reliance on in-class exams. As noted in subpart V(A), this relationship increases the importance of the LSAT as an admissions criterion and correspondingly decreases the likelihood that under-represented minority students will be admitted through the regular admissions process.[186] Although the Court's ruling in *Grutter* permits law schools to look beyond numerical qualifications in order to achieve a critical mass of underrepresented minority students, it can certainly be argued that excessive weighting of the LSAT inflicts positive harm on minority groups even if affirmative action eliminates any adverse admission consequences that flow from the emphasis on speed. In other words, if the typical law school exam is going to emphasize speed of performance, the legal academy needs a justification grounded in educational theory or the testing method's relationship to the work product produced by actual lawyers.[187] In this situation, a rationale

---

writing. This trend is consistent with the relationship between paper grades and UGPA found in the sample as a whole. *See supra* tbls. 9 & 10.

186. *See supra* subpart IV(A); *see also supra* note 32 (citing sources on how differentials in LSAT scores affect minority candidates much more adversely than differentials in UGPA).

187. One practical, rather than theoretical, justification for time-pressured law school exams is that state bar examinations are also conducted under intense time pressure. Although members of the state bar examiners "hierarchy" have had a difficult time defining the bar exam's relationship to competence within the legal profession, one of the relevant skills that the bar exams are supposed to measure is "rapid analysis." Holley & Kleven, *supra* note 32, at 334 & n.129 (discussing the inability and reluctance of bar examiners to articulate a clear theoretical link between the bar exam and the practice of law). In light of the fact that the bar passage rate for minorities has historically been much lower than the passage rate for white applicants, *id.* at 325–28 (providing a detailed review of statistics), the format of the state bar exams may be worth examining. Because of the wide ranging influence of the legal academy, the current format of state bar exams need not be accepted as a given.

grounded in mere tradition may not be enough. Further, it remains an open question whether the Supreme Court's decision in *Grutter* does, in fact, include a 25 year sunset provision.[188] Scrutinizing the issue of speed on law school exams may be a way to make the testing system more meritocratic while achieving the educational benefits of a more diverse student body.[189]

In terms of how time pressure on law school exams may have a disparate impact on minority students, it is important to delineate two separate mechanisms. First, the presence of a significant speededness component on both the LSAT and actual law school exams can increase the predictive validity of the LSAT, thus making it more important for admissions purposes.[190] This linkage has an adverse effect on minority students because the performance gap between white and minority students is larger on the LSAT than on UGPA.[191] The second possible mechanism is that a speededness component has differential effects among particular ethnic subgroups. This is a separate empirical question. As noted in section III(C)(4), the sample size of this study precludes a definitive finding on this issue. However, if future research (entailing a much larger sample size) suggests that time-compressed, in-class exams have a disparate impact on one or more ethnic subgroups, a fairly rigorous theoretical basis will be needed to justify a testing regime in which a large proportion of first-year grades (regional, 80.6%; national, 74.9%) and cumulative law school grades (regional, 74.6%; national, 61.3%) rely on this testing method.

C. *Beyond the LSAT and UGPA: A Preliminary Theory for Predicting Law School Performance*

The central insight of this study is that testing method appears to be a variable that affects the predictive power of both the LSAT and UGPA. In both the national and regional law school samples, the LSAT had its greatest predictive power on in-class exams.[192] This finding is consistent with the hypothesis that a test-taking speed variable is present on both the LSAT and in-class exams. Although UGPA appears to be a more stable predictor than the LSAT as we move between testing methods, the predictive power of

188. *See* sources cited in *supra* note 15 (comparing the language on the 25-year provision used by Justices O'Connor and Thomas with an interpretation advanced by a group of prominent constitutional law scholars).

189. In *Grutter*, the Court ruled that the educational benefit of diversity is a compelling state interest that justified the University of Michigan Law School's admissions policy. Grutter v. Bollinger, 123 S. Ct. 2325, 2339 (2003). I acknowledge that many scholars believe that a diverse student body, particularly at an elite law school, is first and foremost a matter of social justice. I take no position here. My broader point is that a closer examination of law school testing practices may further both the interests of meritocracy and social justice; this situation should be attractive to all factions that participated in the *Grutter* debate.

190. *See supra* subpart V(A).

191. *See supra* note 32 and accompanying text (citing sources on how the differential in LSAT scores affects minority candidates much more adversely than the differential in UGPA).

192. *See supra* tbls. 9, 10, 16, 19, & 21.

UGPA also declines on less time-pressured testing methods.[193] Thus, the results of Phases 1 and 2 suggest that the variable of test-taking speed may explain, at least in part, the uneven predictive power of the LSAT. However, consideration of other variables may be necessary to explain why the predictive power of UGPA also varies by testing method.

A theoretical framework for predicting and interpreting law school performance would likely include several factors. The results of this study suggest at least five: (1) reasoning ability, (2) test-taking speed, (3) motivation and persistence,[194] (4) writing ability, and (5) grading method. Table 25 summarizes how these five factors are relevant to the traditional predictors of law school performance, LSAT, and UGPA.

Table 25: Factors Affecting Performance on LSAT and UGPA

Factor	*Relevant to LSAT*	*Relevant to UGPA*
Reasoning Ability	Yes	Yes
Test-Taking Speed	Yes	Probably
Motivation & Persistence	No	Yes
Writing Ability	No	Yes, varies with curriculum
Objective Grading (i.e., based on identical analytical tasks)	Yes	Yes and No, mixture of papers and exams

Regarding the first two factors, reasoning ability and speed, a student's underlying reasoning ability is probably the most important determinate of that student's LSAT score because it affects a student's ability to identify and select correct answers. However, the number of questions actually reached also affects total score. Thus, if the LSAT is speeded, students with low test-taking speed will tend to have *marginally* lower LSAT scores.[195] A

193. *Id.*

194. Powers has suggested that one of the reasons that the predictive power of the UGPA remains stable, or even increases, between year 1 and years 1–3 of law school is that UGPA may be a measure, at least in part, of motivation and persistence. *See* Powers, *supra* note 88, at 574 (noting that "the contribution of UGPA may increase in later years [of law school] . . . because it better eaptures qualities like persistenee and motivation").

195. As noted in subpart V(A), the presence of a speed component on the LSAT is not likely to be a complete confound. LSAT scores will likely have a strong correlation with the results of an unspeeded test of reasoning ability. *See* KLINE, *supra* note 17, at 72 (noting that there is "a high correlation between the same test [of reasoning ability], timed and untimed"). However, because law school populations tend to enroll students within a fairly narrow LSAT range, students are likely to be relatively *homogeneous* in terms of reasoning ability but relatively *heterogeneous* in terms of test-taking speed. In turn, this heterogeneity has the effect of increasing the predietive ability of the LSAT. This situation would explain Stricker's finding that students with high LSAT

differential in test-taking speed among students could therefore explain the declining predictive power of the LSAT on take-home exams and papers. More specifically, students with high reasoning ability but low test-taking speed will tend to (a) have lower LSAT scores, and (b) perform better (i.e., have a higher ordinal ranking) on take-home exams and papers. The results in section IV(C)(1) support this interpretation. Reasoning ability and test-taking speed (to a lesser degree than the LSAT) are also likely to be reflected in a student's UGPA.

Moving to the third factor, motivation and persistence, this attribute is unlikely to have a significant effect on LSAT scores. The LSAT is designed to measure reasoning ability. Thus, LSAT preparation is generally limited to test-taking techniques and familiarity with format; prep course "gains" are thus benchmarked from what a student would score on the test if he or she took it cold.[196] Further, data on LSAT retests show that gains are generally very modest over time.[197] In contrast, motivation and persistence are very relevant to UGPA because tests at the undergraduate level are designed to measure mastery of subject matter. Regarding the fourth factor, writing ability, insofar as this trait is distinct and separate from reasoning ability, the multiple choice format of the LSAT will not provide an accurate measure. However, depending upon the curriculum, strong writing ability will generally lead to a higher UGPA. Finally, grading criterion is relevant to both the LSAT and UGPA. The LSAT ranks students by comparing performance on identical analytical tasks that can be objectively graded. Because undergraduate testing regimes often require different mixtures of exams and papers, a common objective standard is not always present. Thus, comparisons of UGPA may sometimes be inappropriate or misleading (e.g., science versus liberal arts studies). Table 26 provides a summary of how the five factors affecting LSAT and UGPA are also relevant to four distinctive testing methods found in this study.

---

scores are more prone to "deviantly poor" subscores on the analytical reasoning section of the LSAT. *See* STRICKER, *supra* note 106, at 1. One prominent LSAT study aid singles out the analytical reasoning section as the most time-pressured. *See* ROBINSON & TALLIA, *supra* note 50, at 66 (suggesting that most candidates will only be able to attempt three of the four groups of questions in the analytical reasoning section and that rushing to complete all four is a poor strategy to maximize a candidate's score).

196. *See* Wong, *supra* note 7, at 232 n.157 (noting that "[t]est preparation courses claim that they can boost student test scores by six or seven points" (citing Frances A. McMorris, *Test-Prep Fees Deter Black Law Applicants*, WALL ST. J., Mar. 23, 1998, at B1)).

197. *See, e.g.*, DEBORAH L. SCHNIPKE, LISA C. ANTHONY & LYNDA M. REESE, THE PERFORMANCE OF REPEAT TEST TAKERS ON THE LAW SCHOOL ADMISSION TEST 1 (LSAC, Technical Rep. 98-06, 2000) (providing detailed evidence from five years of LSAT retakes and reporting mean increases of 2.7 points for a second LSAT and 1.5 points for the third time; also noting that the overall mean is consistently highest for first-time test takers because students with low LSAT scores are much more likely to retake the test).

Table 26: Relevance of Five Factors to Law School Testing Methods

Factor	*In-Class Exam*	*Take-Home Exam*	*Paper in Required Legal Writing*	*Paper in Seminar, Elective Course*
Reasoning Ability	Yes	Yes	Yes	Yes
Test-Taking Speed	Yes	Minimal	No	No
Motivation & Persistence	Yes	Yes	Yes	Yes
Writing Ability	Minimal	Moderate	Yes	Yes
Objective Grading (i.e., based on identical analytical tasks)	Yes	Yes	Yes	No (most subjective category)

Starting again with reasoning ability, this attribute is relevant to performance on in-class exams, take-home exams, papers in required legal writing, and seminar papers. Second, test-taking speed will be most important on in-class exams and less important or negligible on the other testing methods. Third, because a law school curriculum involves the mastery of subject matter, motivation and persistence are relevant on all testing methods. Fourth, writing ability is more important for legal writing assignments and seminar papers and less important for law school exams. Correctly identifying and analyzing discrete legal issues are the primary grading criteria of the traditional Blue Book exam; organization and elegant prose are subsidiary factors. Fifth, in-class exams, take-home exams, and legal writing assignments all cover a common intellectual terrain in whieh the instructor has already identified a universe of correct responses. In contrast, seminar papers often cover subject matter chosen by the student. Thus, creativity, depth of analysis, and quality of exposition become the primary grading criteria. This process is obviously more subjective.

As noted earlier, the low predictive power of the LSAT in the national law school sample is probably attributable to range restriction.[198] However, the framework outlined above can be used to explain the changes in predictive ability of the LSAT and UGPA in both the national and regional law school samples. The framework applies as follows.

198. *See supra* note 114 and accompanying text.

1. In-Class Exams.—The correlation between the LSAT and in-class exams is attributable to both reasoning ability and test-taking speed. This correlation is higher in the regional law school sample because of a broader range of ability. Within the national law school sample, the LSAT has its highest predictive power on in-class exams but is poorly correlated with the other testing methods. This pattern suggests that in a range-restricted environment such as the national law school sample, a significant proportion of the predictive power of the LSAT may rest more on differences in test-taking speed than on differences in reasoning ability. In contrast, UGPA is a composite of reasoning ability, test-taking speed, motivation, and persistence. Because all of these attributes are relevant to in-class exams, UGPA predicts fairly well for this testing method.

2. Take-Home Exams.—The LSAT is a less useful predictor of take-home exams in both samples because the variable of test-taking speed is less important. The broader range of ability in the regional law school may explain why the LSAT predicts take-home exams slightly better in the regional sample (0.182) than in the national sample (0.142).[199] UGPA is once again important for reasoning ability and motivation and persistence. The lack of a test-taking speed variable might also have the effect of bunching the distribution of student responses, thus depressing the predictive ability of both the LSAT and UGPA. For example, a professor in the national law school sample gave the same exam to one section using the in-class format and to a second section using the take-home format. In a corresponding exam memo to both sections, the professor commented on his surprise that the in-class exams had a much broader distribution of grades than the take-home exams. This anecdotal evidence corroborates Kissam's assertion that stringent time limits are useful for generating a grading curve with a large number of gradations.[200]

3. Required Legal Writing Assignments.—In both the national and regional law school samples, the first-year legal writing courses are graded on the basis of memos and briefs drawn from identical fact patterns and legal issues. Thus, in addition to organization and writing style, a substantial portion of the legal writing assignments is objective. In other words,

---

199. The time limits on take-home exams were also different in the two samples. The national law school sample typically allotted 8 hours versus 24 hours in the regional law school sample. *See supra* note 126 and accompanying text.

200. Kissam, *supra* note 22, at 453. Although the student population at most law schools is likely to be relatively homogeneous in terms of reasoning ability, it will likely be heterogeneous in terms of test-taking speed. *See supra* notes 112, 195 and accompanying text. Thus, the injection of time pressure is a useful way to generate a curve despite the fact that the resulting student answers tend to be poorly written and organized. *See supra* note 26 and accompanying text.

students must "spot" a discrete universe of issues. Range restriction may explain why the LSAT predicted first-year legal writing grades fairly well in the regional law school sample (0.347) and not at all in the national law school sample (0.036). In contrast, UGPA is a useful predictor of legal writing in both samples because it is a composite of reasoning ability, writing ability, motivation, and persistence.

4. Papers in Seminars and Elective Courses.—In both the national and regional law school samples, the LSAT is a poor predictor of paper assignments in seminars and other elective courses. Although reasoning ability is undoubtedly important to writing lucid and engaging prose, this testing method lacks an objective grading criterion. Thus, students with superior writing skills will tend to do better. A broader range of reasoning skills may account for a modest correlation in the regional law school sample between LSAT and averages in seminar and elective courses (0.124). In the national law school sample, LSAT is uncorrelated with performance on this testing method (0.053). In contrast, UGPA remains a useful predictor of performance on seminar and elective papers because it is a composite of reasoning ability, writing ability, motivation, and persistence.

VI. Conclusion

This study supports the hypothesis that test-taking speed is a variable that affects the ordinal ranking of students on both the LSAT and actual law school exams. The major implication that flows from the data is that the legal academy's heavy reliance on time-pressured law school exams has the effect of increasing the predictive validity of the LSAT. 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. Thus, this same outcome (i.e., ordinal ranking) will tend to replicate itself on time-pressured, in-class exams, thus producing a higher correlation between the LSAT and law school grades. This is not a criticism of the LSAT. Indeed, insofar as law school exams place a substantial—though perhaps inadvertent and under-theorized—emphasis on test-taking speed, the presence of a speededness component on the LSAT may actually make it a better predictor of law school performance. However, on take-home exams and papers, which are arguably more reflective of the systemic time pressure found in the actual practice of law, it appears that the LSAT is actually a weaker predictor of law school performance than UGPA.[201] Therefore, the question that now confronts the legal academy is whether there is a sound

201. *See supra* section IV(B)(3) and tbl.21.

educational, theoretical, or empirical basis for emphasizing speed on law school exams.

Such an inquiry may have substantial payoffs. Understanding the relationship between testing method and the LSAT potentially gives the legal academy a valuable new tool for confronting the LSAT "arms-race" among law schools,[202] while also advancing the dual goals of meritocracy and social justice. When speed is used as a variable on law school exams, the type of testing method, independent of knowledge and preparation, can change the ordering (i.e., relative grade) of individual test-takers. Because the selection of testing method has potentially large distributional consequences for students, legal educators have an obligation to develop a strong theoretical connection between their testing methods and the knowledge and skills that are important in the actual practice of law.[203] Although a migration away from traditional law school "Blue Book" exams may have the socially desirable effect of reducing the importance of the LSAT as an admission criterion, this outcome is largely the by-product of shoring up the theoretical shortcomings that plague the current system of testing.

Finally, this study needs to be kept in context. Despite a rich data set that permitted a wide array of analyses, the sample size is ultimately small and may not be fully representative of other U.S. law schools. Many LSAT studies have encompassed 100 or more schools. What this study gains in depth, it sacrifices in breadth. Yet, notwithstanding the small sample size, the results of this study are sufficiently consistent and robust that they cannot be safely ignored. This study should therefore be replicated and/or expanded to include other law schools. If the evidence is sufficiently compelling, the legal academy will be well-equipped to adopt a more meritocratic and socially just system of law school testing.

202. *See* Wong, *supra* note 7, at 248.

203. In the parlance of psychometrics, law school testing methods should be grounded in a strong "construct theory" that is amendable to empirical testing. *See supra* notes 90–92 and accompanying text.

Appendix 1

Probability Density Functions Plotting Response Time and Accuracy for Questions on Analytical and Logical Reasoning Section of Non-Adaptive Computerized GRE

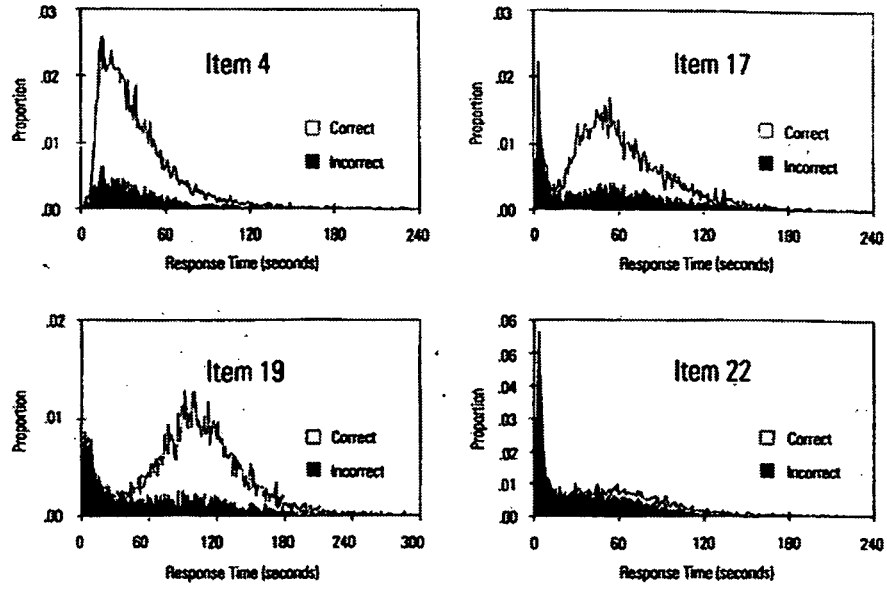


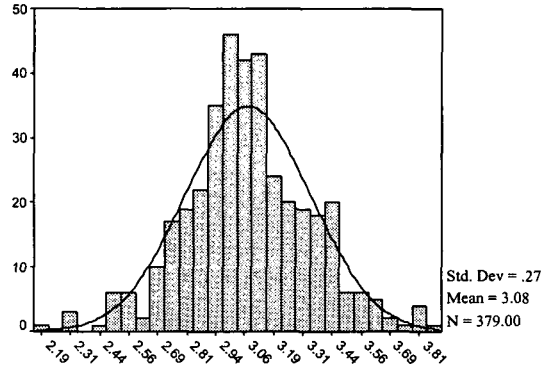
FIGURE 1. *Probability density functions of response times for several representative items. The scales are not the same across items.*

Modeling Response Times

Excerpted from DEBORAH L. SCHNIPKE & DAVID J. SCRAMS, MODELING ITEM RESPONSE TIMES WITH A TWO-STATE MIXTURE MODEL: A NEW APPROACH TO MEASURING SPEEDEDNESS 7 fig.1 (LSAC, Computerized Testing Report 96-02, 1999).

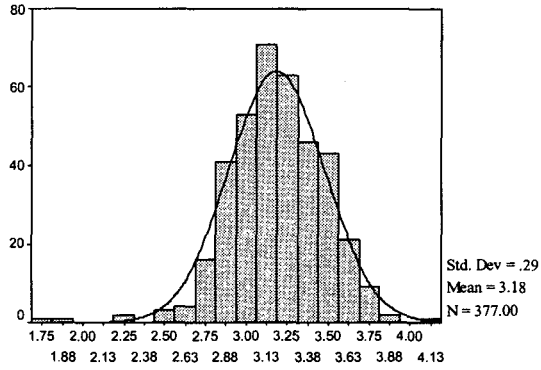
Appendix 2
National Law School

Grade Distributions, In-Class Exams



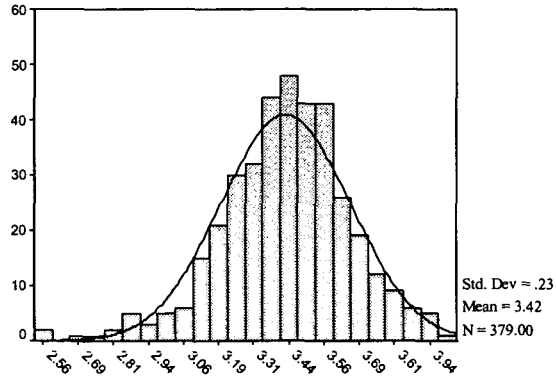
GPA, In-Class

Grade Distributions, Take-Home Exams



GPA, Take-Home

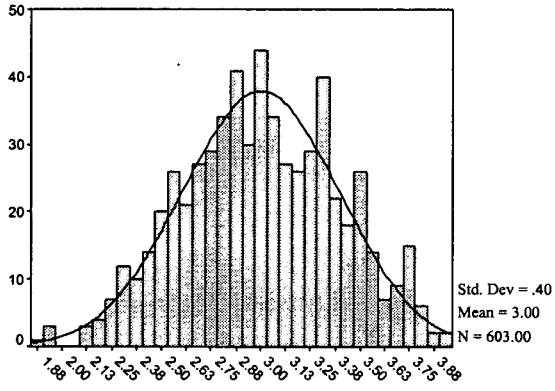
Grade Distributions, Papers



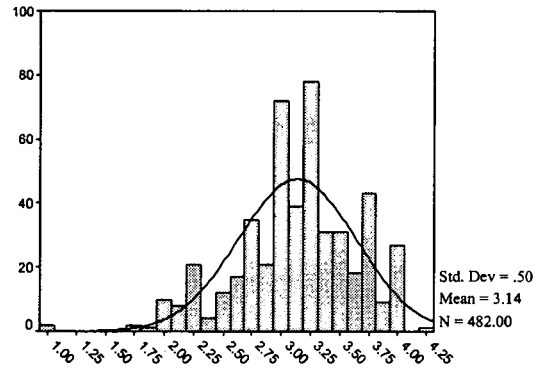
GPA, Paper

Appendix 3
Regional Law School

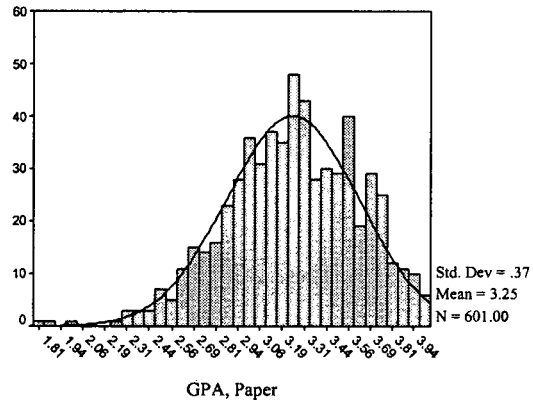
Grade Distributions, In-Class Exams



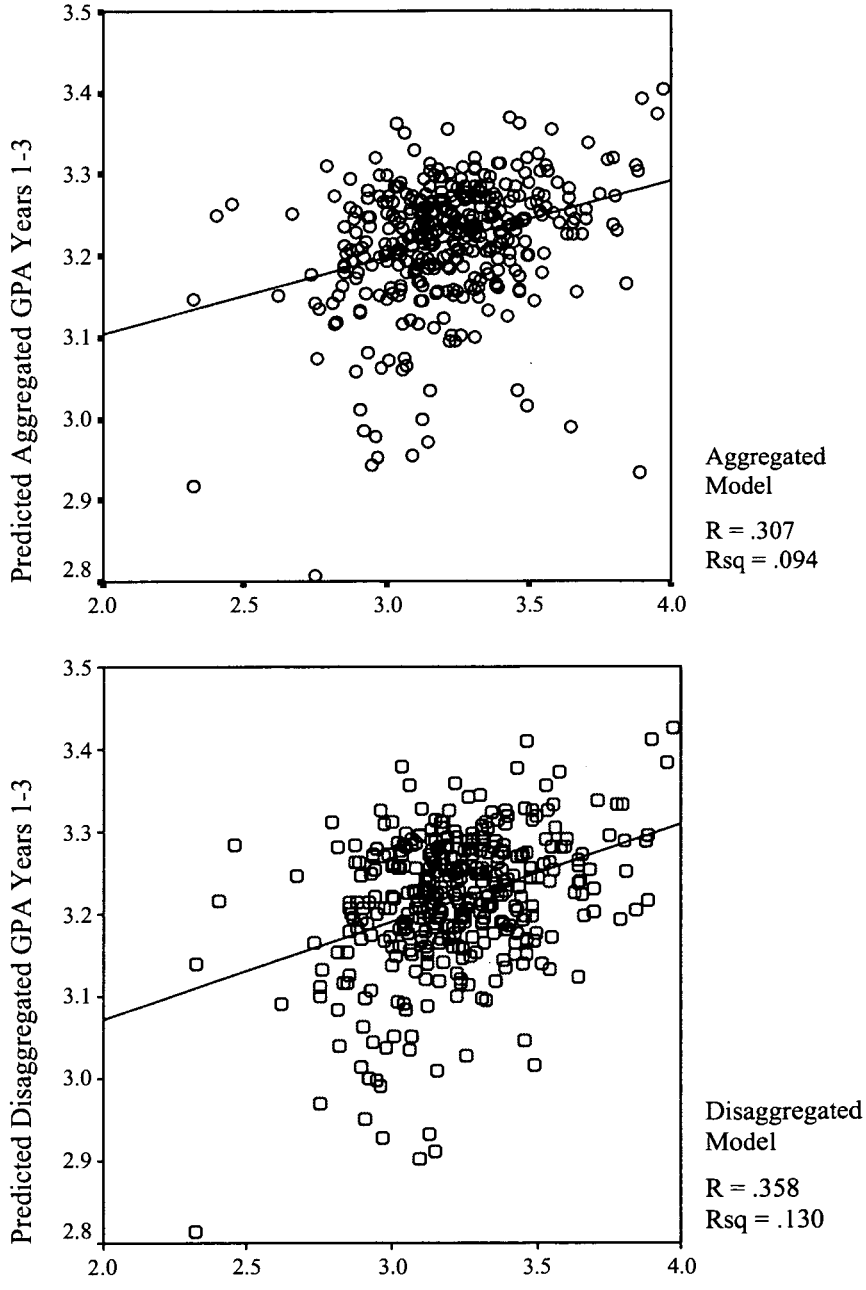
Grade Distributions, Take-Home Exams



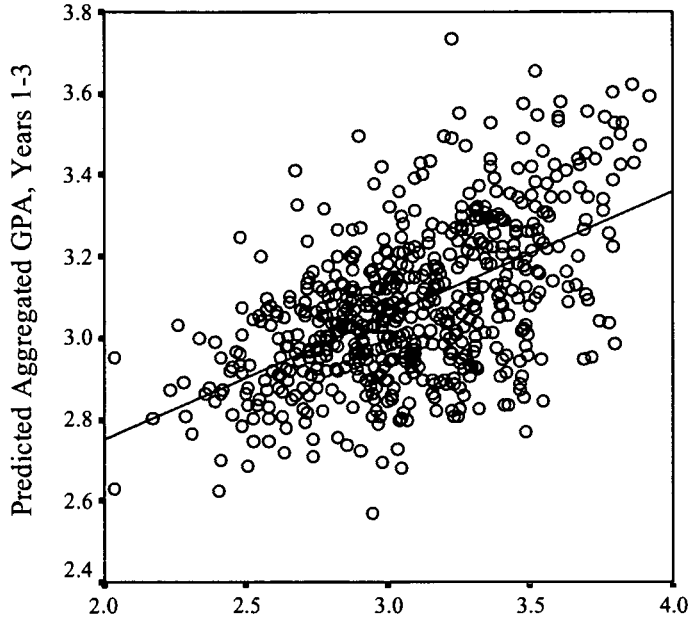
Grade Distributions, Papers

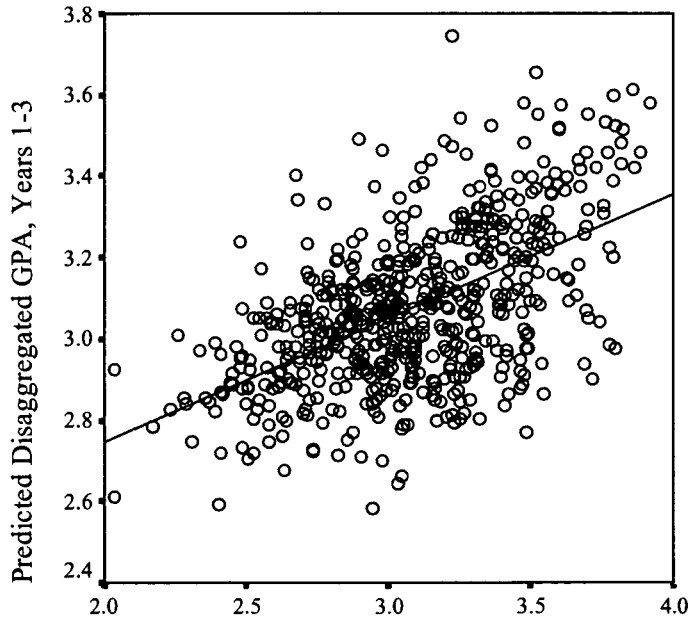


Appendix 4
Scatter Plots for National Law School Sample



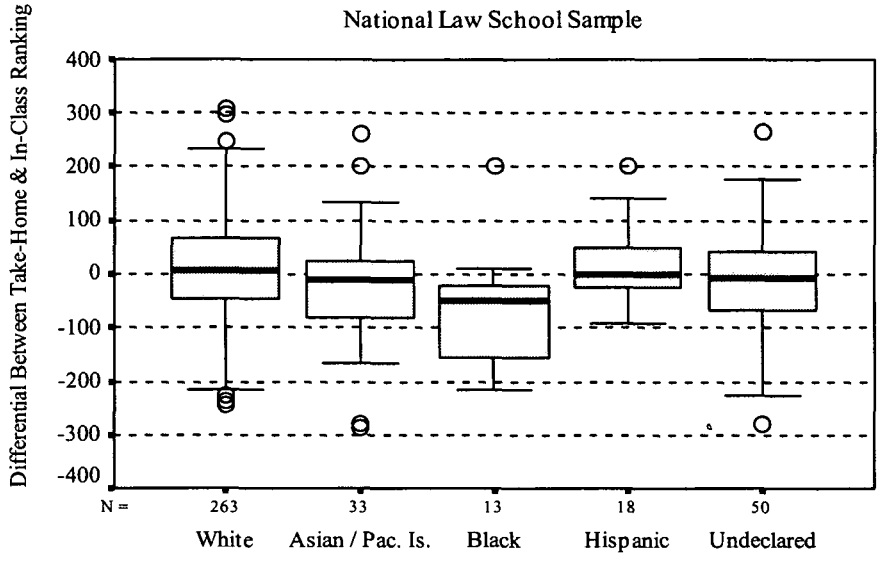
Appendix 5
Scatter Plots for Regional Law School Sample



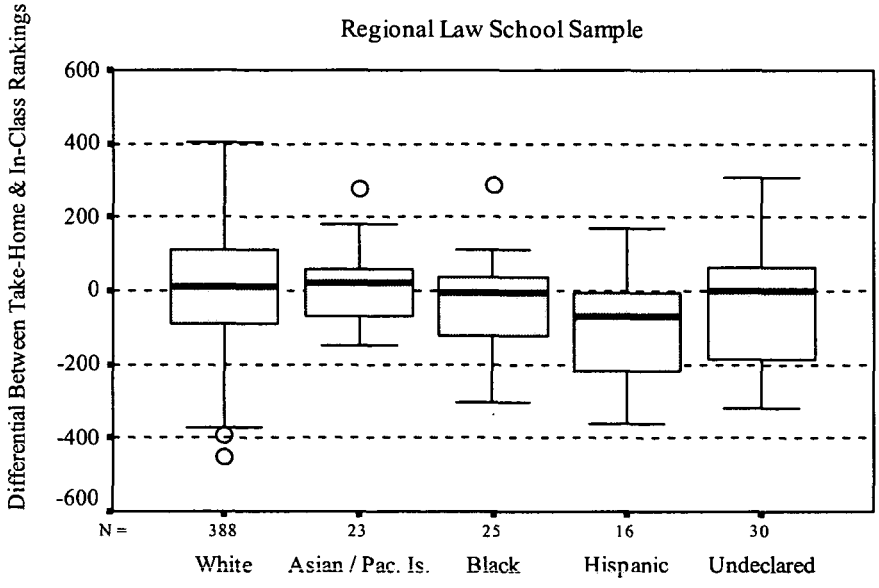


Appendix 6

Change in Ranking between Take-Home & In-Class Exams



Change in Ranking between Take-Home & In-Class Exams



* * *