

ARTICLES

EXTRA TIME AS AN ACCOMMODATION

*Ruth Colker**

Editor's note: *For the benefit of readers interested in particular empirical or psychometric points that this Article explores, Professor Colker provides an Abstract and Table of Contents.*

ABSTRACT

Although the provision of extra time has become the standard method of accommodating students with various disabilities when they take the Law School Admissions Test (LSAT) or in-class timed exams, this Article suggests

* Heck Faust Memorial Chair in Constitutional Law, Michael E. Moritz College of Law, The Ohio State University. I would like to thank Moritz librarian Katherine Hall for her extremely helpful assistance. I would also like to thank my colleague Deborah Merritt who never tired of answering my questions. This Article also benefited from discussions with many members of the Moritz community who engage in admissions decisions; these conversations helped me better understand how the LSAT is used during the admissions process. I would like to thank the participants of a workshop sponsored by the Moritz College of Law on May 23, 2007, for their helpful feedback. I would also like to thank the team of research assistants who helped me with this Article: Matt Johnson, Katherine Klingelhafer, Leah Stoecker and Brad Stoll. Matt Johnson performed the statistical analyses discussed in this paper using SPSS. Research support from the Moritz College of Law and the Ohio State University Distinguished Scholar Award helped make this Article possible. I would also like to thank Anita Bernstein for encouraging me to write this Article for a symposium held at Emory University College of Law. The research conducted for this Article was approved as exempt human subject research by the Institutional Review Board at the Ohio State University. Finally, I would like to thank the University of Pittsburgh and the Thornburgh family for giving me the opportunity to deliver this paper as the Thornburgh Family Lecture on Disability Law and Policy on September 27, 2007.

that we consider other means of attaining fairness on exams. Through a survey of the psychometric literature, as well as a modest empirical study, this Article argues that the rank order of students is likely to be significantly affected by the type of examination instrument used. This Article places the psychometric literature on examination results in the context of testing of prelaw students as well as law students and suggests that we place less emphasis on timed instruments in ranking students.

While questioning the validity of the LSAT for all students, this Article also argues that the LSAT scores of students who take the exam under conditions of extra time are likely to be as valid as the scores of students who take the exam under regular conditions. The Article urges the Law School Admission Council (LSAC) to conduct a proper validity study so that students who receive extra time are not disadvantaged during the admissions process.

Finally, this Article suggests that law faculties examine their testing practices to see if they offer a sufficient variety of testing methods to compare students fairly. The current overemphasis on in-class timed exercises is likely to inappropriately disadvantage certain students and not give them an opportunity to demonstrate their mastery of the material.

Table of Contents

I.	The Case for Extra Time.	419
A.	Processing Speed Evidence.	420
B.	Extra Time as an Accommodation.	421
II.	The LSAT.	425
A.	Development of the LSAT.	425
B.	Scoring of the LSAT.	430
C.	Data on Validity.	433
i.	General Evidence on Validity.	433
a.	Validity of overall results.	433
b.	Validity of discrete test items.	439
c.	Speededness.	442
ii.	Students with Disabilities.	445
a.	The Accommodation Application Process.	445
b.	Validity for Accommodated Test Takers.	450
III.	Law School Exams.	456
A.	The Development of the Law School Exam.	456
B.	Today's Exams.	462
i.	Literature on Law School Exams.	462
ii.	A Modest Survey and Empirical Study of Exams.	465
IV.	Conclusion.	472

INTRODUCTION

The provision of extra time on examinations as an accommodation for some students with disabilities is well accepted. The College Board and the Law School Admission Council (LSAC) have established procedures for students with disabilities to request extra time as an accommodation.¹ Most universities and law schools also have procedures by which such requests can be made.

Litigation has ensued relating to these requests for extra time. In some cases, the plaintiff has contested the denial of a request for extra time.² In other cases, the plaintiff has contested how his or her score is reported when extra time has been provided.³ The issue in those latter cases is whether the score should be “flagged” to indicate that it was taken under accommodated conditions.⁴ In all of these instances, the background presumption has been that extra time is an appropriate accommodation for some students with disabilities.⁵ Similarly, it is presumed that it is appropriate to limit the amount of time provided for examinations so that they contain a “speededness” component.⁶

1. See generally The College Board, SAT Registering with Accommodations, <http://www.collegeboard.com/student/testing/sat/reg/ssd.html> (last visited Aug. 15, 2008); Law School Admission Council, Accommodated Testing, <http://www.lsac.org/LSAT/accommodated-testing.asp> (last visited Aug. 15, 2008) (describing the accommodation process for the LSAT).

2. See, e.g., *Agranoff v. Law Sch. Admission Council*, 17 Nat'l Disability L. Rep. ¶ 223 (D. Mass. 1999) (challenging denial of extra time on LSAT); *Rothberg v. Law Sch. Admission Council*, 28 Nat'l Disability L. Rep. ¶ 129 (10th Cir. 2004) (challenging denial of extra time on LSAT); *Buhendwa v. Univ. of Colo.*, 34 Nat'l Disability L. Rep. ¶ 41 (10th Cir. 2007) (challenging denial of extra time on calculus final exam); *Love v. Law Sch. Admission Council*, 34 Nat'l Disability L. Rep. ¶ 120 (E.D. Pa. 2007) (challenging denial of extra time on LSAT).

3. See, e.g., *Doe v. Nat'l Bd. of Med. Exam'rs*, 199 F.3d 146 (3d Cir. 1999) (challenging flagging of USMLE test results); *Breimhorst v. Educ. Testing Servs.*, No. C-99-338, slip op. (N.D. Cal. Mar. 27, 2000) (challenging flagging of GMAT scores); see also Stephen G. Sireci, *Unlabeling the Disabled: A Perspective on Flagging Scores from Accommodated Test Administrations*, 34 EDUC. RESEARCHER 1, 3 (2004).

4. As will be discussed in Part II.C.ii.b., LSAC takes the position that the test results for test takers who receive extra time have not been “validated” and therefore need to be flagged to reflect that result.

5. See, e.g., *Agranoff*, 17 Nat'l Disability L. Rep. at ¶ 223.

6. The validity of speeded exams is beyond the scope of this Article but deserves significant attention within the law school community. The psychometric literature has long supported the notion that accuracy and speededness are different variables. See generally Susan Ellerin Rindler, *Pitfalls in Assessing Test Speededness*, 16 J. EDUC. MEASUREMENT 261 (1979) (discussing difficulties in measuring speededness of exams); DAVID J. SCRAMS & DEBORAH L. SCHNIPKE, LSAC, MAKING USE OF RESPONSE TIMES IN STANDARDIZED TESTS: ARE ACCURACY AND SPEED MEASURING THE SAME THING? 11 (May,

This Article does not accept those presumptions. Part I assesses the literature that justifies extra time as an accommodation for some students with disabilities. It concludes that the provision of extra time as an accommodation is more controversial than is frequently recognized in the legal literature. Part II applies these observations to the Law School Admissions Test (LSAT), and Part III applies these observations to law school examinations. Part IV concludes by asking what approaches—other than providing extra time—we might use in conducting fair exams.

As this Article argues, the issue of extra time on examinations is very complicated. Undoubtedly, certain disabilities cause some people to perform language-processing tasks more slowly than others. In other words, it may take some individuals longer to demonstrate their knowledge and understanding in a testing situation because of a disability. If these individuals with disabilities are given an insufficient amount of time to complete an exam, then we have tested for their “disability” rather than their “ability.” For example, two students may be given a “reading check” to see if they have understood material they read for homework on the previous night. If student *A* completes the test and student *B* does not, we do not know whether student *B* in fact knew the answers to the questions not completed. If student *B* did not complete the test due to a disability which impaired her processing speed, then we have measured her disability (i.e., her slow processing speed) but have not fully measured her understanding of the assigned reading. Under these testing conditions, we can compare the speed of students *A* and *B*, but we cannot compare their mastery of the material.

A typical solution to this problem is to give the individual with a disability additional time to take the examination so that she can demonstrate her ability. The complication is that individuals with disabilities are not the only ones who might better demonstrate their abilities with extra time. In my previous hypothetical, let’s assume that the test was an essay test, and the students were asked to write about the material they read the previous evening. If the testing conditions are timed, student *B* might not be able to answer all

1999), available at <http://www.lsacnet.org/Research/Making-Use-of-Response-times-in-Standardized-Tests-Accuracy-Speed-Measuring-Same-Thing.pdf> (“For the present data, speed and accuracy are at least separable if not orthogonal.”); Jens Forster et al., *Speed/accuracy Decisions in Task Performance: Built-in Trade-Off or Separate Strategic Concerns?*, 90 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 148 (2003) (stating that speed/accuracy trade-offs may be influenced by the strategic inclinations of the participants); DEBORAH L. SCHNIPKE & DAVID J. SCRAMS, LSAC, EXPLORING ISSUES OF TEST-TAKER BEHAVIOR: INSIGHTS GAINED FROM RESPONSE-TIME ANALYSES (Mar. 1999), available at <http://www.lsacnet.org/Research/Exploring-Issues-Test-Taker-Behavior-Insights-Gained-Response-Time-Analyses.pdf>.

the questions and would certainly benefit from the provision of extra time. Even though student *A* was able to complete the exam in the standard time, student *A* might also be able to improve the quality of his answer if provided extra time. In fact, the empirical literature suggests that *most* test takers improve their performance when offered extra time.⁷ A further complication is that many tests are given to compare test takers' performances; the results are curved. The LSAT and law school exams share that central feature.⁸ If extra time is to be given to some test takers in a comparative, curved context, then it is important for the extra time to be apportioned *fairly*, especially if all test takers are likely to improve performance with extra time. How much extra time is appropriate? How would that determination be made?

This Article offers some insight into how to think about some of these questions. It suggests that our current approach to extra-time allocation is haphazard at best. Alternatively, we could seek to lessen the "speededness" component of exams so that these basic issues of fairness arise less starkly. Unfortunately, it is not easy to create a law school exam or standardized testing instrument which is "unspeeded." Creating fair testing instruments for all students is more challenging than others have previously acknowledged. The difficulty of the exercise, however, should not cause us to stop trying to attain fairness. Fairness must be our governing principle even if our practices fall short of that goal.

Part I of this Article examines the literature that supports the provision of extra time for some students with disabilities. It argues that the "differential boost" theory, which is often used to justify extra time for some students with disabilities, is more complicated to apply in practice than is often recognized. Part II applies the literature on timed exams to the LSAT. It discusses the history of the development of the LSAT and shows how it has emerged as a speeded exam that is currently used on a rank-order basis to predict first-year grades. It argues that the LSAT is not a particularly strong predictor of first-year grades, but it is probably an equally good predictor of first-year grades for students who receive extra time and for those who do not. More importantly, Part II questions why the LSAT takes its current form as a highly "speeded" exam that is scored on a rank-order basis. This type of examination has a disparate impact against students with undiagnosed learning disabilities

7. See *infra* Part I.B.

8. The bar exam may be distinguishable. One could view the bar exam as a competency exam rather than a comparative, scaled exam. If so, the application of this Article to that context would be different than to the LSAT or law school exams.

as well as students who excel in accuracy but not speed (and who might make very good lawyers).

Part III discusses law school exams, suggesting that they, too, often include an overemphasis on speed rather than accuracy. Part IV concludes that a focus on students with disabilities can help us see the unfairness of the LSAT and law school exams for many students, including those who do not receive accommodations. Rather than try to make these exams fairer through the accommodation process, this Article urges LSAC and law professors to try to make these exams fairer for all students by decreasing the emphasis on speededness.

I. THE CASE FOR EXTRA TIME

The argument for extra time as an accommodation is based on two observations: (1) certain disabilities cause students to proceed more slowly when taking written examinations; and (2) the provision of extra time “evens the playing field” so that these students can demonstrate their abilities—rather than their disabilities—when taking an exam. The first observation is well established; the second is more controversial.

This Article focuses on requests for extra time by students diagnosed with learning disabilities (LD)⁹ or Attention Deficit Disorder (ADD)¹⁰ because

9. The term “learning disability” has various definitions. See Kenneth A. Kavale & Steven R. Forness, *What Definitions of Disability Say and Don't Say: A Critical Analysis*, 33 J. LEARNING DISABILITIES 239 (2000). These researchers argue that the most frequently used definition of learning disability is the legal definition found in the Individuals with Disabilities Education Act. *Id.* at 240. These researchers suggest a more sophisticated, operational definition that tries to identify more precisely the nature of the student's psychological process deficit. *Id.* at 251. The diagnosis of a learning disability can be challenging because the tests used to measure aptitude might, themselves, have a component which is problematic to a student with a learning disability. Hence, researchers have extensively debated how to measure the existence of a learning disability. See generally David J. Francis et al., *Defining Learning and Language Disabilities: Conceptual and Psychometric Issues with the Use of IQ Tests*, 27 LANGUAGE, SPEECH & HEARING SERVICES IN SCH. 132 (1996); David J. Francis et al., *Psychometric Approaches to the Identification of LD: IQ and Achievement Scores are Not Sufficient*, 38 J. LEARNING DISABILITIES 98 (2005); Kenneth A. Kavale et al., *Responsiveness to Intervention and the Identification of Specific Learning Disability: A Critique and Alternative Proposal*, 28 LEARNING DISABILITY Q. 2 (2005); Briley Proctor & Frances Prevatt, *Agreement Among Four Models Used for Diagnosing Learning Disabilities*, 36 J. LEARNING DISABILITIES 459 (2003). In this Article, I presume that certain students have measurable learning disabilities, and I do not engage the literature on how to determine the existence of a learning disability other than to discuss some problems with the methodology used by LSAC to determine the existence of a learning disability.

10. See generally RUSSELL A. BARKLEY, *ATTENTION-DEFICIT HYPERACTIVITY DISORDER: A HANDBOOK FOR DIAGNOSIS AND TREATMENT* (3d ed. 2006). For a discussion of the relationship between ADHD and learning disabilities, see Laurie E. Cutting & Martha Bridge Denckla, *Attention: Relationships*

those are the two categories that present the most requests for extra time as an accommodation; they are also the categories for which the most empirical research is available. Although these results may be generalizable to other kinds of disabilities, this Article makes no claim for such generalizability. Given the conflicting nature of the research concerning LD and ADD, it seems important to proceed cautiously before making assumptions about the validity or appropriateness of accommodations.

A. Processing Speed Evidence

The educational psychology literature consistently establishes that children with LD¹¹ in reading have slower processing speeds than other children.¹² A recent study by Michael Weiler and a team of researchers confirmed this finding.¹³ Their research documented that children with LD score within the expected age range “on measures of Oral Language, Motor Speed, and Visual Spatial skill, but less well . . . on measures of Written Language, consistent with their well-documented phonological processing problems.”¹⁴

The evidence about students with ADD is more complicated. Weiler and his researchers focused their attention on children with attention deficit disorders who were impulsive and inattentive, but not hyperactive.¹⁵ They

Between Attention-Deficit Hyperactivity Disorder and Learning Disabilities, in HANDBOOK OF LEARNING DISABILITIES 125 (H. Lee Swanson et al. eds., 2003).

11. See, e.g., M. Kay Runyan, *The Effect of Extra Time on Reading Comprehension Scores for University Students with and Without Learning Disabilities*, 24 J. LEARNING DISABILITIES 104 (1991) (finding that students with learning disabilities performed as well as other students on a reading test when both groups were given additional time but performed less well under time-constrained conditions).

12. This Article focuses on students who have a learning disability that produces slower processing times in reading. Not all students with learning disabilities, however, have difficulties with reading. Instead, some could have a deficit in language, writing, or math. See Kavale & Forness, *supra* note 9, at 252. There is also some dispute between a specific learning disability in reading and a reading disability. See Kavale et al., *supra* note 9, at 4. This Article does not engage that dispute. Most requests for accommodation on standardized tests are from students who claim to have a learning disability in reading. See Lawrence J. Lewandowski et al., *Assessment of Reading Rate in Postsecondary Students*, 21 J. PSYCHOEDUCATIONAL ASSESSMENT 134, 135 (2003).

13. Michael D. Weiler et al., *Processing Speed in Children with Attention Deficit/Hyperactivity Disorder, Inattentive Type*, 6 CHILD NEUROPSYCHOLOGY 218 (2000).

14. *Id.* at 227. One problem with these studies as applied to law students is that they are typically performed on children. There is no consensus on how to assess reading rates on adult populations. See Lewandowski et al., *supra* note 12, at 134 (reporting that the curriculum-based assessment method is a more accurate testing instrument than the Denney Reading Rate measure for adult students even though the Denney is the more commonly used instrument).

15. Weiler et al., *supra* note 13, at 219.

concluded that these students were similar to students with LD. They experienced “diminished speed of processing.”¹⁶ From this study, they concluded that educators should consider “reducing the pace at which information is delivered” and “relaxing time constraints” for this group of students, because the source of their problems may be “efficiency of cognitive processing.”¹⁷

The Weiler study is one of many studies to conclude that children with slower processing speeds should be provided extra time as an accommodation for their disability.¹⁸ As the next section demonstrates, however, the leap from slower processing speed to extra time is more complicated than acknowledged by some researchers.

B. Extra Time as an Accommodation

From the prior discussion, it is clear that a disability makes it difficult for some students to demonstrate their knowledge under speeded circumstances because they may simply not have enough time. We could give extra time to the student with a disability, but then what do we do with the other students? Should they receive extra time as well?

The conventional answer, under the “differential boost” theory, is no.¹⁹ Under this theory, students with disabilities would benefit *more* from a particular accommodation than other students.²⁰ Nondisabled students are able to demonstrate their knowledge and ability through the regular exam process; their performance is not significantly affected by the provision of an accommodation.²¹ By contrast, students with a disability would receive a

16. *Id.* at 229.

17. *Id.* at 230.

18. See, e.g., Jennifer Hartwig Lindstrom, *The Role of Extended Time on the SAT for Students with Learning Disabilities and/or Attention-Deficit/Hyperactivity Disorder*, 22 LEARNING DISABILITIES RES. & PRAC. 85 (2007) (concluding that SAT results can be interpreted in the same way when students with disabilities have an extended-time administration as compared to the standard-time administration); Nicole Ofiesh et al., *Using Speeded Cognitive, Reading, and Academic Measures to Determine the Need for Extended Test Time Among University Students with Learning Disabilities*, 23 J. PSYCHOEDUCATIONAL ASSESSMENT 35 (2005) (acknowledging the widespread use of extra time on exams and suggesting which tests are appropriate for determining allocation of extra time).

19. See Lynn S. Fuchs & Douglas Fuchs, *Helping Teachers Formulate Sound Test Accommodation Decisions for Students with Learning Disabilities*, 16 LEARNING DISABILITIES RES. & PRAC. 174, 175 (2001).

20. *Id.* at 175-76.

21. *Id.* at 176.

differential boost through the provision of the accommodation.²² An example would be the provision of a large type-face exam. A student with normal vision would experience no change in test performance through the use of an exam with large type-face. A student with a visual impairment, however, might experience a marked increase in performance.

The differential boost theory becomes more difficult to justify in the context of an accommodation like extra time because that accommodation might benefit both students with and without disabilities. Further, whether extra time produces a differential boost might depend on the type of examination given. A reading exam might produce different results than a math exam when extra time is allocated.

Although extra time is a frequently provided accommodation for the LSAT and law school exams, few studies have validated that accommodation under the differential boost hypothesis. Studies have been conducted of both math and reading tests.²³ On tests of basic math skills, studies have not supported the differential boost theory.²⁴ Students with and without learning disabilities achieved about the same improvement in performance with extended time.²⁵ The exception to this pattern was a study in which students were asked to complete complex math problems.²⁶ Extended time particularly benefited students who had an underlying learning disability in reading but no learning disability with respect to math.²⁷

Not all studies of reading comprehension have validated the differential boost theory. A study of 400 students, 200 with and 200 without learning disabilities, was unable to confirm the differential boost theory.²⁸ Students without learning disabilities benefited at least as much from extended time as students with learning disabilities.²⁹ By contrast, M. Kay Runyan was able to validate the differential boost theory in her study of university students.³⁰ She administered a reading test to thirty-one students, sixteen of whom were

22. *Id.*

23. *Id.* at 175-76.

24. *Id.* at 176.

25. *Id.*

26. *Id.*

27. *Id.* at 176; see also Stephen N. Elliott & Ann M. Marquart, *Extended Time as a Testing Accommodation: Its Effects and Perceived Consequences*, 70 *EXCEPTIONAL CHILDREN* 349 (2004) (reporting that students without disabilities benefited at least as much with extra time as students with disabilities on a math test).

28. See Fuchs & Fuchs, *supra* note 19, at 175.

29. *Id.* at 175-76.

30. See Runyan, *supra* note 11.

identified as having learning disabilities.³¹ The students were given twenty minutes to take the test; they were then allowed to use whatever time was needed to complete the test if they had not completed it within twenty minutes.³² For the normally achieving students, the provision of extra time made no significant difference in performance.³³ For students with learning disabilities, the provision of extra time made a significant difference in performance.³⁴ Moreover, the extra time “evened the playing field” by causing the students with learning disabilities to have test scores that were not significantly different from the normally achieving students.³⁵ Without accommodations, however, their scores were significantly different.³⁶ Studies by Nicole Ofiesh also support the differential boost theory,³⁷ although G.E. Zuriff argues that methodological problems distort the findings from some studies that purport to support the differential boost theory.³⁸

A recent study examined the differential boost theory for students with Attention Deficit Hyperactivity Disorder (ADHD) in the context of a math exam.³⁹ The researchers concluded that students without ADHD actually benefited more from extra time than students with ADHD.⁴⁰ The students with ADHD were typically taking medication for their condition.⁴¹ They

31. *Id.* at 105.

32. *Id.* at 106.

33. *Id.* at 106, tbls.1-2.

34. *Id.*

35. *Id.* at 107.

36. *Id.* at 106.

37. With a sample size of thirty university students with learning disabilities and thirty university students without learning disabilities, Ofiesh found that the LD students, in general, received more benefit from extended time than non-LD students. See Nicole S. Ofiesh, *Using Processing Speed Tests to Predict the Benefit of Extended Test Time for University Students with Learning Disabilities*, 14 J. POSTSECONDARY EDUC. & DISABILITY 39 (2000). In a more recent study, Nicole Ofiesh, Nancy Mather, and Andrea Russell, also conducted research that supports the differential boost theory. See Ofiesh et al., *supra* note 18.

38. See G.E. Zuriff, *Extra Time for Students with Learning Disabilities: An Examination of the Maximum Potential Thesis*, 13 APPLIED MEASUREMENT EDUC. 99 (2000). Zuriff critiques five studies, none cited in this Article, which purport to support the differential boost (or what he calls the maximum potential) thesis. His criticism of small sample size and research design problems would arguably apply to the Runyan and Ofiesh studies cited in this Article. The Fuchs study, which used a much larger sample, does not support the differential boost theory, as discussed above. Although it is true that students with learning disabilities have slower processing speeds as a result of their impairment, the Zuriff critique should make us aware that it is challenging to find appropriate responses to recognition of that disability.

39. See Lawrence J. Lewandowski et al., *Extended Time Accommodations and the Mathematics Performance of Students with and Without ADHD*, 25 J. PSYCHOEDUCATIONAL ASSESSMENT 17, 25 (2007).

40. *Id.* at 25.

41. *Id.* at 21.

examined the number of correct answers on a math test after twelve minutes and after eighteen minutes for a group with ADHD and a control group without ADHD.⁴² The students were between ten and thirteen years of age.⁴³ Although both groups benefitted from extended time, the researchers found that the *nondisabled* students benefitted *more* from extra time than the students with ADHD.⁴⁴ Interestingly, they found that the performance of the two groups did not differ significantly if the students with ADHD were given time-and-a-half and the control was given standard time.⁴⁵

What can we learn from these varied results? At a minimum, they reflect how difficult it is to determine the appropriate amount of extra time under a differential boost theory.⁴⁶ In the Runyan study, for example, it appears that the normally achieving test takers had an ample amount of time to take the test through the provision of twenty minutes. Given the extra time, to them the test already felt like it was “untimed.” Hence, the availability of extra time did little to boost their performance. The students with learning disabilities, however, did not typically complete the test within twenty minutes.⁴⁷ Hence, extra time differentially boosted their performance. In the other studies discussed above, the normally achieving students appear to have felt like the test was “speeded” so they benefited from the provision of extra time. To see a differential boost effect—to minimize the effect of extra time on the normally achieving students—the researchers would probably have had to allow much more time for all study participants to complete the test under regular conditions.

These results present a quandary for law school test administrators. If normally achieving law students find the LSAT or law school exams to be “speeded,” is it fair to give extra time to students with various disabilities? If so, how do we determine how much extra time provides a “speeded” equivalence?

In the law school context, one significant challenge is that each professor gives a different exam for each test administration. As a result, we cannot use historical data to predict how much extra time would provide a “speeded”

42. *Id.* at 22.

43. *Id.* at 20.

44. *Id.* at 25.

45. *Id.*

46. For an excellent article discussing the challenges involved in assessing the appropriate amount of extra time is appropriate, see Nicole S. Ofiesh & Charles A. Hughes, *How Much Time?: A Review of the Literature on Extended Test Time for Postsecondary Students with Learning Disabilities*, 16 J. POSTSECONDARY EDUC. & DISABILITY 2 (2002).

47. *Id.* at 5 tbl.2.

equivalence. Further, the number of students who receive accommodations is so small that it is difficult even to conduct an empirical analysis. LSAC has a less significant challenge because of the large number of test takers and the comparatively large number of students who take the exam under conditions of accommodation. The appropriate amount of extra time probably has to be individualized for each student. However, it is doubtful that LSAC can make a precise decision for each student irrespective of how much data it collects from physicians and others about test takers.

None of these caveats mean that we should give up trying to provide extra time to some students with disabilities on exams. The empirical evidence is clear: Some students do have disabilities that slow down their processing speed, and these students can only demonstrate mastery of a subject matter with more than the standard time allocated for exams. In the next two sections, I explore what LSAC and law schools have done to deal with the “speededness” of their exams. In the final section, I offer recommendations in light of these findings.

II. THE LSAT

A. *Development of the LSAT*

The first record of testing for admissions purposes was at Columbia Law School in 1921, when Dean Harlan Stone initiated an experimental testing program.⁴⁸ Dean Merton Ferson proposed a law school admission test in the 1920s to law school deans with the cooperation of West Publishing Company and the assistance of psychologist George Stoddard.⁴⁹ It was administered to current first-year law students to determine whether it was a reliable predictor of their performance.⁵⁰ That test was used for several years by some law schools.⁵¹ It was called the Stoddard-Ferson Aptitude Test.⁵² Yale Law School began aptitude testing for admissions purposes in the mid-1920s, the University of California developed an admissions test in 1938, and the University of Iowa developed a legal aptitude test in 1943.⁵³

48. Thomas O. White, *LSAC/LSAS: A Brief History*, 34 J. LEGAL EDUC. 369, 369 (1984).

49. Merton L. Ferson, *Law Aptitude Examinations*, 5 AM. L. SCH. REV. 563, 564 (1926).

50. *Id.*

51. White, *supra* note 48, at 369.

52. *Id.*

53. *Id.* at 369-70.

Law school admissions were very different at that time than they are today. Students did not necessarily have to obtain a four-year degree before entering law school, and the number of applicants was not nearly as large as it is today. For example, Yale Law School sought to admit a class of 120 students out of approximately 400 applicants.⁵⁴ Some argued that it was better to use an aptitude test to weed out students who would not pass their first year of law school than to use first-year grades for the weeding out process.⁵⁵

These early tests had their critics, including Professor John H. Wigmore.⁵⁶ Wigmore was not opposed to the concept of testing, but he claimed that his data from Northwestern Law School suggested that the Stoddard-Ferson Test was not a good predictor of first-quarter grades or grades for the entire law school experience.⁵⁷ One of the test developers observed, in reply, that the Stoddard-Ferson Aptitude test was only to be given to applicants who “have already met various standard criteria of admission to a law school”⁵⁸ and was to be only one of many criteria in law school admissions.⁵⁹ Henry B. Witham, Acting Dean, College of Law, University of Tennessee, conducted a study to assess the predictive validity of the Stoddard-Ferson Aptitude test at his law school.⁶⁰ He concluded that the test was best at predicting who would be in the bottom one-quarter of the law school class.⁶¹ Wigmore focused his attention on the validity of the instrument for the top quartile of the class and had done a very rudimentary analysis of only eleven students in a class of sixty-seven.⁶² In fact, a fuller study of his entire data set reveals that the aptitude test was reasonably predictive of grades during law school.⁶³ Ironically, these early tests offered some of the highest predictive validity of any tests administered to law school applicants.

54. Albert B. Crawford & Tom Jay Gorham, *The Yale Legal Aptitude Test*, 49 YALE L.J. 1237 (1940).

55. *See id.*

56. *See* John H. Wigmore, *Juristic Psychopoyemetrology—Or, How to Find Out Whether a Boy Has the Makings of a Lawyer*, 24 ILL. L. REV. 454 (1930).

57. *Id.* at 463. In fact, an analysis of this data suggests that the test did a decent job of predicting first year grades—as good or better than the current LSAT. The Pearson Correlation between AB Record and law school admission test was 0.495, and between A Record and LSAT was 0.470. By contrast, as we will see in Part II.C., the correlation between law school grades and LSAT scores is around 0.41.

58. George D. Stoddard, *Correspondence: Legal Aptitude Tests*, 25 U. ILL. L. REV. 446, 447 (1931).

59. *Id.*

60. *Id.* at 448.

61. Henry B. Witham, *Correspondence: Legal Aptitude Tests*, 25 U. ILL. L. REV. 448, 450 (1931) (concluding that 75% of those who scored in the bottom one-fourth of the test also placed in the bottom one-fourth of their law school class).

62. Wigmore, *supra* note 56, at 461.

63. *See supra* note 57.

Yale Law School devised its own aptitude test—a ninety-minute exam consisting of verbal comprehension, logical inference and analogies, and legal material.⁶⁴ Initially, the test had a predictive correlation coefficient of 0.64 of predicting first-year grades in law school, which is considered good for that kind of testing instrument.⁶⁵ Because of the importance of factors such as hard work and the variability of grading among professors, one wouldn't expect any factors to be able to predict law school grades with absolute precision. Hence, social scientists typically consider a predictive index above 0.50 as *good* and above 0.70 as *excellent*.⁶⁶

These early ventures into testing caused the Association of American Law Schools to express interest in a general examination to be given to all applicants to law schools.⁶⁷ Although this interest was first expressed in 1938, serious efforts did not begin until after World War II ended. Most “law schools had reduced operations to a bare minimum” during World War II, so interest in testing lay dormant during that period.⁶⁸

The first version of what is now the LSAT was developed in 1947.⁶⁹ That year also marked a time when law schools found themselves faced with a growing applicant pool as many World War II veterans sought to apply to

64. Crawford & Gorham, *supra* note 54, at 1238. The content of this test appears similar to various modern tests except for one section in which the student is expected to recall the facts of a case five minutes after being given an opportunity to study it. *Id.* at 1239.

65. *Id.* at 1241.

66. *Id.* at 1241 n.6. Assessing the validity of such a testing instrument through a predictive validity study of first-year grades poses many difficulties which are discussed further in Part III.B of this Article. Two are worth mentioning with respect to these early attempts at validation. First, such a study only examines those who attend law school. We have no way of knowing the law-school performance of candidates who were denied admission. Second, the range of performance of those who are studied is relatively narrow. At Yale, for example, more than three-fourths of all grades were between 70 and 79. *Id.* at 1245. Crawford and Gorham used those factors as explanations for why the predictive validity coefficient fell from 0.64 to 0.55 during the course of their study. They said: “Therefore, it follows necessarily that the more any factor (whether college grades or test scores) is actually utilized in selecting students, so as to curtail their distribution by progressively cutting off the lower end, the less likely will that factor correlate with the criterion it purports to predict.” *Id.* at 1241. Nonetheless, even their 0.55 coefficient is good within the standards of the testing community. (It was also better than the 0.46 and 0.38 values found for the Ferson-Stoddard and the Thorndike tests used at the University of Illinois and Columbia Law School, *id.* at 1248, and better, we will see, than the LSAT.) They also argued that a primary purpose of the test was to weed out failures and that the failure rate did decline following the use of this test as one part of the admissions process. *Id.* at 1242.

67. White, *supra* note 48, at 370.

68. *Id.*

69. William P. LaPiana, *Merit and Diversity: The Origins of the Law School Admissions Test*, 48 ST. LOUIS U. L.J. 955, 963 (2004).

college and law school.⁷⁰ Initially, law schools hoped to use the LSAT as one factor among many in admitting applicants and, principally, to weed out unqualified applicants rather than to use scores on a rank-order basis.⁷¹ In the early years, it appears that the test did not produce a wide range of scores so that it could not be used on a rank-order basis.⁷² Hence, suggestions were made about how to increase the difficulty of the test to produce a wider range of scores.⁷³

The LSAT was not initially developed to play the central role that it currently plays in the admissions process and law school rankings. When Yale first began offering an admissions test in 1930, it administered the test at the second stage of a two-step process.⁷⁴ When the LSAC decided to develop its own LSAT examination, it actually debated whether to allow candidates to know their results on the test.⁷⁵ Had the test designers envisioned the importance the test would play in the admissions process, it is hard to imagine that they would have considered not telling the candidates their scores. Today, of course, not only do candidates know their LSAT scores, but law schools themselves know their median LSAT score, which plays a crucial role in the *U.S. News & World Report* ranking system.⁷⁶

The most consistent theme with respect to the content of the LSAT is that it has undergone extensive change over time. The first version of the LSAT was administered in 1948 and consisted of ten sections administered over the course of an entire day.⁷⁷ The content of the various subtests changed nearly

70. *Id.* at 962.

71. *Id.* at 977.

72. Louis Toepfer of Harvard Law School apparently complained in 1956 that everyone had high scores. *Id.* at 978.

73. *Id.*

74. *Id.* at 956-57.

75. *Id.* at 970-71.

76. The LSAT score is one of the two most significant factors in the *U.S. News* ranking system. *See id.* at 988. The “LSAT scores a weight of 12.5% in calculating each law school’s ranking” in *U.S. News and World Report*. David A. Thomas, *Predicting Law School Academic Performance from LSAT Scores and Undergraduate Grade Point Averages: A Comprehensive Study*, 35 ARIZ. ST. L.J. 1007, 1008 (2003). Further, according to a study commissioned by the Association of American Law Schools, “90% of the overall differences in ranks among schools can be explained solely by the median LSAT score of their entering classes.” *See* Williams D. Henderson & Andrew P. Morriss, *Student Quality as Measured by LSAT Scores: Migration Patterns in the U.S. News Rankings Era*, 81 IND. L.J. 163, 165 (2006). This factor has caused one commentator to report that there is an “arms race” among law schools for higher LSAT scores. *See id.* at 166.

77. LYNDA M. REESE & RUTH ANNE COTTER, LSAC, A COMPENDIUM OF LSAT AND LSAC-SPONSORED ITEM TYPES 1948-1994, at 1 (Law Sch. Admission Council Research Report 94-01, 1994), available at <http://www.lsacnet.org/Research/Compendium-of-LSAT-and-LSAC-Sponsored-Item-Types.pdf>.

every testing administration; the entire test was shortened to a half day in 1951.⁷⁸ The content of the subtests continued to change until 1961, when an afternoon session consisting of a writing ability and general background section was added.⁷⁹ The afternoon session continued until 1971, when it was dropped.⁸⁰ The content of the subtests continued to change. From 1947 to 1984, LSAC experimented with forty-one different subtests on the LSAT.⁸¹ Arguments for inclusion or exclusion tended to focus on the predictive validity of, as well as difficulties with grading or scoring, a given subsection.

In addition to the changes in its content, the length of the test has also varied. It was a six-hour test from 1948 to 1951.⁸² It then became a three-and-a-half-hour test.⁸³ In 1961, it became a full day, three-score test, with scores on general background, writing, and LSAT.⁸⁴ The afternoon session was eliminated in 1971; the test became three hours and fifty minutes.⁸⁵ In 1974, the test was reduced to three hours and twenty minutes.⁸⁶ Presently, the test is three hours and twenty-five minutes.⁸⁷

Today's law professors would probably be surprised at the content of the current version of the LSAT. It contains three scored sections: Logical Reasoning, Analytical Reasoning, and Reading Comprehension.⁸⁸ There is also a non-scored Writing Sample.⁸⁹ It would be difficult to do well on this exam without studying for the specific types of questions. In the Logical Reasoning section, the test taker has to be able to identify the type of reasoning used.⁹⁰ Test takers who prepare for these questions can learn skimming strategies such as eliminating answers with non-matching

78. *Id.* at 2.

79. *Id.* at 2.

80. *Id.*

81. *Id.* at 3-5 tbl.1.

82. *Id.* at 6 tbl.2.

83. *Id.*

84. *Id.*

85. *Id.*

86. *Id.* at 7 tbl.2.

87. *Id.*

88. KAPLAN TEST PREP & ADMISSIONS, KAPLAN PREMIER PROGRAM 4 (2007).

89. *Id.*

90. The hardest part of this section appears to be the "Parallel Reasoning" section. The test taker is required to determine which answer choice uses reasoning similar to the given argument. This section requires the test taker to analyze six arguments to find the best match. Alternatively, the test taker is asked to answer the choice that is flawed in the same way as the given argument. Again, the test taker needs to analyze six arguments to find the best match. Finally, some questions in this section ask the test taker to identify which answer choice best describes the *method* of argument used by the author. If the test taker does not know the definition of the methods identified, this question is very difficult.

conclusions; otherwise, this section can require time-consuming and careful reading that is hard to muster in the time permitted.

The hardest section for a student who has not studied for the exam would probably be the Analytical Reasoning section. It consists of logic games. As various study guides explain, there are shortcuts that one can employ to do logic games, such as drawing a sketch of the information provided.⁹¹ This section would seem to be easiest for students with strong nonverbal aptitude who can “see” the logic game. Anecdotally, I have been told that these questions are extremely difficult for visually impaired students who cannot readily draw a picture of the rules from the logic game. The labeling of these sections as “Logical Reasoning” and “Analytical Reasoning” gives them apparent facial validity, but closer examination suggests that students could be both analytical and logical, yet not be sufficiently test-savvy to do well on these questions.⁹²

The Reading Comprehension section is most similar to questions that students have probably seen on other standardized exams. They read passages and respond to questions about those passages. Studying can probably improve those skills, but unlike the other two parts of the exam, the format of the exam questions in this section is probably familiar to most students.

In sum, the LSAT is not an exam that a student might casually take and expect to do well on. Two-thirds of the test consists of time-pressured, detailed questions, preparation for which would benefit the test taker.

B. Scoring of the LSAT

The LSAT is a timed testing instrument. At the present time, the LSAT consists of four scored thirty-five minute sections and one unscored thirty-five minute section, which is used for future “norming” purposes.⁹³ The sections average around twenty-five questions so that students, in total, answer

91. KAPLAN, *supra* note 88, at 88.

92. I sent some logical reasoning questions to my faculty and asked them to tell me what answers they were able to record within four minutes (the time allotted for those questions). Most faculty were able to provide me with the correct answers but most also reported they had to use more than the allotted time to get these answers. The logic game question took several minutes to answer which caused them to exceed their time allotment. Some of the younger members of my faculty, who had helped teach in LSAT prep courses, were able to answer the questions more quickly because they knew some short cuts. This was certainly not a scientific sample but it was interesting to learn that, although we place significant weight on LSAT scores during the admissions process, many faculty might do poorly on the LSAT.

93. KAPLAN, *supra* note 88, at 4.

between 101 and 103 scored questions in 140 minutes.⁹⁴ In addition, there is a thirty minute unscored writing section.⁹⁵ With the inclusion of the unscored experimental section and the writing section, the total test time is three hours and twenty-five minutes. The results from the multiple choice questions are then reported on a scaled score between 120 and 180.⁹⁶ In other words, there are sixty different outcomes that are possible on a test with about 100 questions. There is no penalty for guessing; 99.2% of the test takers scored above a 128—which corresponds to 22-23 correct answers—on the June 2007 administration of the LSAT.⁹⁷ Hence, about 77 or 78 raw score points are used to divide up students on a 52-point scale.⁹⁸

The test is curved so that five or six scaled score points are used at the top of the scale to distinguish among the top one percent of the test takers.⁹⁹ In other words, a scaled score of 174 corresponds to a raw score of 95 whereas a score of 180 corresponds to a raw score of 99-100. Yet, a 174 is the 98.9 percentile and a 180 is the 99.9 percentile. Because of the curve, the scores are quite bunched around the mean. The mean score (51.9 percentile) is 153, which corresponds to a raw score of 65-66. One or two more correct answers increase the student's scaled score to 154 and the 56.2 percentile.¹⁰⁰ An equivalent 4.3% leap at the top of the scale would require a student to answer as many as ten more answers correctly. (The LSAT score of 168 corresponds to the 95.5% which, in turn, corresponds to a raw score of 90.)¹⁰¹ In other words, small changes in the number of correct answers around the mean result in large percentage movements up or down the scale; scale changes in the number of correct answers result in little percentile movement at the top (or very bottom) of the scale.¹⁰²

94. *See id.*

95. *See id.*

96. *See Id.*

97. *See* Law Sch. Admission Council, Form 8LSN75, at 38 (2007), *available at* <http://www.lsac.org/pdfs/2007-2008/SamplePTJune07.pdf> (Conversion Chart) [hereinafter LSAC Test Form].

98. *Id.*

99. *Id.*

100. *Id.*

101. *Id.*

102. William Kidder makes this point in his article on the gender bias inherent in use of the LSAT, using data from an earlier administration of the LSAT:

The median score on the LSAT is a 151. Starting with an example of two students, suppose that Michelle obtained a 150 on a recent LSAT and Michael scored a 152 on the same test. In practical terms, this means that Michael answered only two more questions correctly out of the 101 scored questions on the test. In the high-stakes contest of law school admissions, these two questions will place Michael at the 54th percentile of applicants and Michelle at the 46th percentile. Because there have been as many as 100,000 law school applicants in recent years,

It is important to understand how the LSAT is scored because it provides a better sense of the importance of each test question. If a student's goal is to receive a perfect, 180 score, the student can achieve that score with a raw score ranging from 98 to 101.¹⁰³ However, if a student's goal is to score in the top one percentile, then the student can achieve that result with a raw score ranging from 88 to 101—a 13-point range.¹⁰⁴ If a student takes the test for the first time and scores at the 50th percentile (score of 151) and has a goal of raising that score to the 60th percentile (score of 153), the student only needs to answer correctly four or five more questions.¹⁰⁵

These small differences are important because of the rank-order way that law schools use LSAT scores. In fact, the LSAT has a significant “error of measurement” that is inherent in each score. At a ninety-five percent confidence interval, the error of measurement is 5.2 points.¹⁰⁶ Hence, scores of 151 and 153 are statistically indistinguishable, yet the difference between them is likely to have a significant impact on a candidate's ability to be admitted to a law school.¹⁰⁷

For the average test taker—the candidate who scores around 151—small changes in the number of correct answers can have a big impact on his or her score. In turn, whether he or she has enough time to answer the questions could have a big impact on his or her performance. LSAC has not done direct studies of the time-pressured aspect of the exam for all test takers, but its studies of accommodated students give some insight into the time-pressured elements of the test. When students with disabilities take the exam twice—once without accommodations and once with accommodations—their scores typically improve by more than eight scaled points, which can be equivalent to a twenty-percentile improvement.¹⁰⁸ LSAC reports that test takers who were accommodated with extra time indicated that they had

the two additional questions that Michelle missed can rank her several thousand places behind Michael in the national applicant pool.

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, 18-19 (2000) (footnotes omitted).

103. See LSAC Test Form, *supra* note 97, at 38.

104. *Id.*

105. *Id.*

106. Kidder, *supra* note 102, at 19.

107. LSAC does report scores in a “band” to suggest that schools do not give undue weight to small differences in test scores.

108. ANDREA THORNTON ET AL., ACCOMMODATED TEST TAKER TRENDS AND PERFORMANCES FOR THE JUNE 1993 THROUGH FEBRUARY 1998 LSAT ADMINISTRATIONS, at 13-14 (Law Sch. Admission Council Technical Report 01-04, 2001), available at <http://www.lsacnet.org/research/Accommodated-Test-Taker-Trends-and-Performance.pdf> [hereinafter THORNTON ET AL., ACCOMMODATED TEST TAKER TRENDS].

enough time (but not too much time) to complete the multiple choice section of the exam, but LSAC provides no comparative data for non-accommodated test takers.¹⁰⁹ Some regular test takers may have ample time to complete the exam, but for those who do not, the scores may reflect their lack of speed rather than their ability with respect to some of the responses. If accommodated test takers are given an opportunity to take the test with ample time, then it would seem appropriate for all test takers to have the same opportunity.

C. Data on Validity

i. General Evidence on Validity

a. Validity of overall results

The concept of “validity” is used to explain what inferences are permissible from test results.¹¹⁰ The LSAT is used as a screening device for law school admissions. There are two general ways that one could validate the LSAT for that purpose: construct or predictive validity.¹¹¹

By “construct validity,” we mean that “the consistency in test responses and score correlates to some quality, attribute, or trait of persons or other objects of measurement.”¹¹² If we used construct validity for the LSAT, we might argue that it directly measures the skills and abilities necessary to be a good lawyer. LSAC has never attempted to validate the LSAT through a construct validity model, presumably because it is nearly impossible to define the skills and abilities necessary to be a good lawyer.¹¹³ Even if we could define those skills, it would certainly be challenging to design a test that could measure them. Skill inventories for lawyers have included such diverse abilities as collecting facts, interviewing clients, legal analysis, oral presentation, written exposition, and law office mechanics.¹¹⁴

109. *Id.* at 9.

110. SAMUEL MESSICK, *VALIDITY OF TEST INTERPRETATION AND USE* 2 (1990) (“Broadly speaking, validity is an inductive summary of both the existing evidence for and the actual as well as potential consequences of score interpretation and use. Hence, what is to be validated is not the test or observation device as such but the inferences derived from test scores or other indicators.”).

111. See generally PAUL KLINE, *A PSYCHOMETRIC PRIMER* (2000); MESSICK, *supra* note 110.

112. MESSICK, *supra* note 110, at 3-4.

113. See generally H. Russell Cort & Jack L. Sammons, *The Search for “Good Lawyering”: A Concept and Model of Lawyering Competencies*, 29 CLEV. ST. L. REV. 397 (1980).

114. *Id.* at 401 n.20.

“Predictive validity” models, by contrast, do not purport to directly measure the characteristics needed to be successful in law practice. Instead, “[p]redictive validity indicates the extent to which an individual’s future level on the criterion is predicted from prior test performance.”¹¹⁵ The criterion could “be performance as a lawyer.” In that case, we would ask how the LSAT score predicts future performance as a lawyer. But it is virtually impossible to define “performance as a lawyer.” So instead, LSAC seeks to validate the LSAT by saying it can help predict first-year grades in law school.¹¹⁶ LSAC has never tried to validate the LSAT under the construct validity model; instead, it asserts that the LSAT can be used to help predict performance during law school.¹¹⁷

The fact that LSAC does not seek to validate the LSAT under a construct validity model can easily be misunderstood by a lay person. Because the content of the LSAT looks a bit like material that a student might see in law school, one might think that LSAC is seeking to directly measure skills needed in law school. In fact, LSAC even claims that the test “is designed to measure skills considered essential for success in law school.”¹¹⁸ But, in reality, that

115. MESSICK, *supra* note 110, at 7.

116. In a predictive validity study, a correlation coefficient is generated which ranges from -1.0 to 1.0 and describes the relationship between the score and the predicted criterion (in this case, performance in law school). If the correlation coefficient were 1.0 then we would say that a perfect relationship existed between the test score and the predicted criterion. In real life, a perfect score of 1.0 is highly unusual.

If we want to know how much explanatory information the test provides, there is a common method to determine that information based on the observed coefficient. The measure of the variance in the criterion that is explained by the predictor is defined by squaring the coefficient. A correlation coefficient of [0].3, for example, means that the test explains only 9% of the variation in predicted performance. In other words, the test leaves unexplained 91% of the variance reflected in the performance measure.

Michael Selmi, *Testing for Equality: Merit, Efficiency, and the Affirmative Action Debate*, 42 UCLA L. REV. 1251, 1263-64 (1995) (footnote omitted).

117. LSAC’s registration information for the LSAT states,

The LSAT is designed to measure skills considered essential for success in law school: the reading and comprehension of complex texts with accuracy and insight; the organization and management of information and the ability to draw reasonable inferences from it; the ability to think critically; and the analysis and evaluation of the reasoning and arguments of others.

LAW SCH. ADMISSION COUNCIL, LSAT & LSDAS INFORMATION BOOK 1 (2007), available at <http://www.lsac.com/pdfs/2007-2008/Infobooktext2007web.pdf> (last visited Mar. 9, 2007). This statement by LSAC actually combines both criterion and predictive validity measures because it asserts what specific skills are needed to perform well during law school. In fact, LSAC doesn’t have any data proving that the test really does test for those skills; it simply knows that the test has some ability to predict performance during the first year of law school. For purposes of this Article, it is also interesting to note that LSAC does not assert that speed is one of the qualities needed to perform well during law school; although, as I will discuss, the test is a strong measure of speed.

118. *Id.*

is not an accurate way to understand LSAC's attempts to validate the testing instrument. If it was attempting to validate the test through construct validity, it would need to have psychologists closely study those who do particularly well in law school (or even as lawyers) and try to create a test that would measure the skills and abilities of those people. Construct validity is comparatively easy for jobs that require physical skills. If a truck driver needs to be able to drive a truck safely, then we might give truck-driver applicants a road test in which they demonstrate their skills. Giving an adequate road test might be time consuming, so we might prefer to give a paper-and-pencil test to a roomful of applicant truck drivers. We might find, for example, that people who become successful truck drivers do well on a test that examines how to change the oil on an automobile. They may not need that skill as a truck driver, yet the acquisition of that skill correlates well with the skill that they do need.

The problem with a predictive validity model is that all sorts of extraneous or unacceptable factors might predict success in law school. For example, having a family member who is a lawyer might correlate with success in law school, so might socio-economic status, but we would probably not consider those to be acceptable criteria for admission. One's score on a Sudoku puzzle or a crossword puzzle might also correlate with success, but those criteria might also be considered too arbitrary to be used. So instead, LSAC administers a test that looks like it might be valid under a construct validity model but actually validates it under a predictive validity model.

Predictive validity measures, however, do not offer strong evidence of reliability. A predictive validity coefficient for predicting something like grades is rarely higher than 0.5, which means it predicts no more than 25% of the variance in the outcome measure.¹¹⁹ LSAC uses a predictive validity model to claim that the LSAT is a good predictor of first-year grades in law school. LSAC could probably use a test that was heavily weighted towards reading comprehension and get about the same predictive validity as if it chose a test weighted heavily towards mathematics. The first test might select a heavily female population and the second test might select a heavily male population, but both might be equally valid under a predictive validity model. Hence, if several different types of test items produced the same predictive validity, then there is some arbitrariness in which test questions are actually used. LSAC has never been able to develop a testing instrument with higher than a 0.41 predictive validity coefficient, but it has developed *many* different

119. Selmi, *supra* note 116, at 1263-64.

testing instruments that achieve about that outcome.¹²⁰ Each of these testing instruments does not select the *same* population; they each select populations equally likely to do well in law school. Therefore, the fact that a testing instrument meets a benchmark score as a predictive validity measure does not mean that it could not be modified to be equally valid yet select a different population.¹²¹

Although LSAC has been historically satisfied with its correlation coefficient of 0.41 for predictive validity, not all educational psychologists would be comfortable with that figure. Professor M.A. Brimer's critique of correlation coefficients of 0.70 underscores the weakness of such claims:

We can assume that we have constructed an examination which has a mean of 50 and a standard deviation of 10. We can then look at the case of a candidate who obtains a mark of 50. The mark of such a candidate would be subject to such a degree of error that we could have only minimal confidence that his true mark would fall somewhere between 15 and 85. In 5% of cases our candidates with an obtained mark of 50 would have true marks outside even these limits. When one remembers the arguments that go on over the difference between as little as 5 marks in an examination, the enormity of our practice in making such discriminations is apparent.¹²²

His argument suggests that it is inappropriate to use an exam on a rank-order basis with predictive validity values of 0.41.

Given that predictive validity is the dominant tactic used to validate the LSAT, one should be aware of the inherent limitations of such a strategy. In a predictive validity study, an applicant's LSAT score is compared to the applicant's first-year grades in law school.¹²³ The sample only includes those who actually attend law school. Thus, although the average LSAT score for all test takers is 151, the average LSAT score for those whose data is used in a predictive validity study is 156.¹²⁴

120. For the discussion on the correlations between LSAT scores and first-year law school grades, see STILLWELL ET AL., *infra* note 130 and accompanying text.

121. With respect to this Article, an interesting question is whether the LSAT's predictive validity would change if *all* applicants were given 50% more time to complete the exam but it otherwise remained unchanged.

122. M.A. Brimer, *Professional Examinations*, 10 J. SOC'Y PUB. TCHRS. L. 37, 43 (1969).

123. ANDREA E. THORNTON ET AL., LSAC, PREDICTIVE VALIDITY OF ACCOMMODATED LSAT SCORES, at 6 tbl.7 (Law Sch. Admission Council Technical Report 01-01, 2001), available at <http://www.lsacnet.org/research/Predictive-Validity-of-Accommodated-LSAT-Scores.pdf> [hereinafter THORNTON ET AL., PREDICTIVE VALIDITY].

124. *Id.* at 8 tbl.7.

From the outset, attempts were made to validate the LSAT in a predictive validity model. In 1945, Frank H. Bowles, admissions director at Columbia Law School, sought to develop an examination to assist law school admissions.¹²⁵ He asked the College Board to develop a test that would have a “correlation coefficient of 0.70 or higher.”¹²⁶ By 1947, it was agreed that a test would be developed to be used by many of the leading law schools, and they would seek to have the results “correlate with first-year grades.”¹²⁷ By this measure, the test was considered to be successful from the outset because it was reported to correlate well with first-year grades, although the exact correlation was unclear.¹²⁸ A 0.70 correlation would be excellent because it would suggest that the LSAT could predict 49% of the variation in law school grades.¹²⁹ In fact, the LSAT’s validity does not approach the 0.70 benchmark specified when the test was first developed. In a 2002 report, researchers for LSAC report that the predictive validity of the LSAT, when used alone, to predict first year grades is 0.41, meaning that it predicts about 16% of the variation in first-year grades.¹³⁰

The 0.41 correlation coefficient, which has been static over time, is far worse than the goal set by LSAC when the test was first developed. As LSAC notes, there are at least two measurement reasons for that imprecision. First, “there is a certain amount of measurement error inherent in the test.”¹³¹ Schools tend to admit students within a relatively small band of LSAT scores. In real terms, there may not be much difference between students at the top and bottom of that range when we consider the five-point error band associated with each individual score. The second, and related, reason is that the selected students within a school tend to be “more homogenous than the entire applicant pool.”¹³² This is the “restricted pool” problem.

The small correlation coefficient should caution us from using the LSAT on a rank-order basis. The fact that the correlation coefficient is highest for

125. LaPiana, *supra* note 69, at 962.

126. *Id.*

127. *Id.* at 963.

128. *Id.* at 976.

129. “An indication of how much agreement there is between sets of scores may be obtained by squaring the correlation coefficient. Thus a correlation of 0.8 indicates 64% agreement.” KLINE, *supra* note 111, at 25.

130. See LISA ANTHONY STILWELL ET AL., PREDICTIVE VALIDITY OF THE LSAT: A NATIONAL SUMMARY OF THE 2001-2002 CORRELATION STUDIES, at 5 tbl.2, 11 tbl.6 (Law Sch. Admission Council Technical Report 03-01, 2005), available at <http://www.lsacnet.org/Research/Predictive-Validity-of-the-LSAT-National-Summary-of-the-2001-2002-Correlation-Studies.pdf>.

131. *Id.* at 4.

132. *Id.*

those in the top one-quarter of the pool suggests that the LSAT may be a useful mechanism for predicting who might be the most successful, although, even for that group, the correlation coefficient is only 0.46,¹³³ suggesting that it accounts for about 21% of the variance. By contrast, the LSAT is a poor predictor of law school performance for those in the bottom quartile of the exam. For the bottom quartile, the correlation coefficient is only 0.36 for 2001 and 0.34 for 2002,¹³⁴ suggesting it predicts about 12% of the variance for those test takers. Although early studies of law school admissions tests suggested that the test was particularly good at weeding out people who would otherwise have difficulty earning passing marks in law school, the current version of the LSAT does not seem to serve that purpose well.¹³⁵

It is true that the predictive validity of the LSAT improves to .50 when used in conjunction with undergraduate grade-point average,¹³⁶ meaning that it can predict 25% of the variance in first-year grades. That combined measure is commonly called the “index” and can assist schools in making admissions decisions.

But the most well-known national ranking publication, *U.S. News & World Report*, does not rank on the basis of the index.¹³⁷ It separately reports

133. *Id.* at 5 tbl.2.

134. *Id.*

135. This is a complicated assertion, however, because very few students currently flunk out of law school. The test does serve a “weeding out” function by causing students not to be admitted to various law schools, and we have no way of knowing how students would have performed if they had been admitted. One might argue that the fact that few students flunk out shows that the LSAT is serving its weeding out function well. But the fact that the LSAT does not accurately predict who will be in the bottom quartile of the class should at least make us suspect that many factors other than abilities measured on the LSAT predict poor performance in law school.

136. See STILWELL ET AL., *supra* note 130, at 5 tbl.2. Although LSAC apparently studied the predictive validity of various subtests, items were retained despite poor evidence of reliability. The “Practical Judgment” test item fell into that category. It was used from 1975-1982. REESE & COTTER, *supra* note 77, at 5 tbl.1, 14-15. Despite the fact that predictive validity studies showed it had “no predictive value,” it was used for eight years. *Id.* at 15. I still remember taking this test item when I took the LSAT in 1978. Until I figured out a trick for doing this type of question, it totally baffled me. You had to indicate whether an item was a major objective, major factor, minor factor, major assumption, or unimportant issue. *Id.* at 70-71. Those terms were poorly defined. The trick to getting the right answer was to construct an outline. A minor factor was a sub-point on the outline. The word “minor” didn’t mean less important; it meant tertiary. I wasn’t surprised to see that this test item did not correlate at all to first year grades in law school. It probably correlated with whether someone had studied the test questions in advance. I’m sure that I would have done very poorly on this section if I hadn’t looked at the types of questions in advance. The fact that I can remember this one section of the exam nearly thirty years later suggests that it struck me as very arbitrary at the time. LSAC’s own data revealed that it was an arbitrary test component, yet they used it for eight years.

137. See U.S. NEWS & WORLD REPORT, AMERICA’S BEST GRADUATE SCHOOLS 45 (2008 ed.).

schools' median LSAT score and median GPA. It then gives more weight to the LSAT than the GPA in its ranking system. Further, it does not produce a combined score for each student that takes into account both his or her LSAT and GPA. Because women, on average, perform less well than men on the LSAT, but perform as well as men in law school,¹³⁸ this ranking system disadvantages schools that have admitted women based on their projected law school grades, despite their lower LSAT scores.

One conclusion that can be drawn from this extensive discussion is that the LSAT is not a strong predictor of success in law school. Its validity, however, will depend, in part, on how law school grades are determined. If the LSAT, for example, would be given under conditions of "extreme speed" and law school exams were also given under such conditions, then we would expect to see a strong correlation even if neither instrument was truly a valid measure of what is required to be a competent lawyer. But this relationship is complicated by the overall low predictive reliability of the testing instrument. One problem with an "extreme speed" testing instrument is that there may be an excessive number of students ranked at the bottom because they do not have time to do anything more than guess the right answer. Those students might be at the bottom of both the LSAT and law school exams, but neither exam would do a very good job of distinguishing among them. A testing instrument is more likely to produce a high validity coefficient if everyone has an opportunity to demonstrate their knowledge, because then there will not be a cluster of "guessers." On the other hand, if a testing instrument is quite easy, the untimed version might produce an undifferentiated cluster at the top. The challenge is to find a testing instrument that has little or no speededness component and produces a range of responses under non-speeded circumstances.

b. Validity of discrete test items

LSAC studies have also examined the characteristics of discrete test items rather than overall predictive validity of the examination taken as a whole. Some test items were dropped because they were hard to construct or expensive to grade. For example, the "Debate" test item produced relatively high predictive validity scores (between 0.44 and 0.75) but was abandoned because it "was difficult to construct."¹³⁹

138. See *infra* note 216.

139. REESE & COTTER, *supra* note 77, at 10-11.

Because LSAC is constantly experimenting with the LSAT, its content has varied considerably over time. From 1976 to 1982, for example, it was mathematical in orientation because it included a section called “Quantitative Comparison,” in which “the test taker was required to compare two quantities and state which is greater, state if they are equal, or state that there is insufficient information to make a judgment.”¹⁴⁰ Sample test questions required the test taker to use principles of geometry to calculate angles or lengths of sides.¹⁴¹ Not surprisingly, some studies showed that this section favored white males.¹⁴² Nonetheless, this test item was as “valid” as other test items which favored reading skills because validity was defined in terms of predicting first-year grades, not using the skills necessary to be a lawyer.

LSAC has often experimented with using writing sections on the LSAT even though those sections have never been found to correlate well with success in the first year of law school. The use of these items reflects the tension between a test that is assessed on the basis of predictive validity and a test that is assessed based on construct validity. Writing skills are considered important to lawyers yet do not appear to predict success in law school. A test of writing ability was used from 1961 to 1970 with a separate score being given for that part of the test.¹⁴³ “[V]alidity studies showed that the writing measures pretested did not contribute to the prediction of writing ability and did not improve the prediction power of the LSAT.”¹⁴⁴ A predictive validity perspective allows LSAC to justify instruments that emphasize quantitative skills because such skills will help predict first-year grades in law school. If the first year of law school, however, were transformed to look more like the practice of law, with an emphasis on polished writing, one must wonder what kind of content would have to be contained in the LSAT for it to predict first-year grades.

LSAC’s conclusion that a writing test does not predict first-year grades is consistent with a finding by William Henderson that LSAT scores do a poor job of predicting grades on writing projects during law school.¹⁴⁵ At a highly selective national law school, where students frequently wrote papers and had high incoming LSAT scores, he found that the predictive correlation

140. *Id.* at 16.

141. *See id.* at 75 (Quantitative Comparison, Part A).

142. *Id.* at 16.

143. *Id.* at 22.

144. *Id.*

145. WILLIAM D. HENDERSON, SPEED AS A VARIABLE ON THE LSAT AND LAW SCHOOL EXAMS 1 (Law Sch. Admission Council Research Report 03-03, 2004), available at <http://www.lsacnet.org/research/Speed-as-a-Variable-on-the-LSAT-and-Law-School-Exams.pdf>.

coefficient from the LSAT to in-class exams was 0.265 but dropped to 0.057 for assigned papers.¹⁴⁶ At a regional law school, where there were fewer assigned papers but a broader range of LSAT scores, he found that the validity coefficient dropped from 0.453 for in-class exams to 0.285 for assigned papers.¹⁴⁷ Those differences are statistically significant.¹⁴⁸ It is interesting that most, if not all, law schools require students to write papers during law school, presumably because they think that the skills developed by writing papers are important to the practice of law, yet they use a testing instrument for admission to law school that does not predict whether students will perform well on untimed written exercises like papers. Although not perfect, undergraduate GPA was a better predictor of performance on assigned papers than LSAT scores in Henderson's study.¹⁴⁹

Henderson's data has some inherent limitations because he does not control for whether the paper courses use the same curve as the exam courses.¹⁵⁰ At most law schools, courses with papers have smaller enrollment than courses with exams and tend to have a higher overall grade distribution. The restriction in grade range makes it more difficult to have a correlation between LSAT scores and grades in writing courses. Henderson did find that the LSAT score correlated better with the grade in a paper course when the paper course was a first-year legal writing course.¹⁵¹ It is common for first-year legal writing courses to be graded on an established curve that models the other first-year courses, which could help explain the higher correlation with the LSAT score. Nonetheless, the correlation coefficient between grades in a first-year required legal writing course at a regional law school and LSAT scores was lower than between grades in courses with in-class exams and LSAT scores,¹⁵² which suggests that the LSAT does a comparatively poor job of evaluating whether students can engage in high quality legal writing.

146. *Id.* at 12 tbl.9.

147. *Id.* at 12 tbl.10.

148. *Id.* at 13 tbls.11, 13.

149. For the national law school sample, the correlation coefficient for UGPA was 0.192 for assigned papers as compared with a correlation coefficient of 0.057 for the LSAT. For the regional law school sample, the correlation coefficient for UGPA was 0.333 for assigned papers as compared with 0.285 for the LSAT. Henderson doesn't report whether those differences are statistically significant. *Id.* at 12 tbls.9, 10.

150. Henderson's data reflect that paper courses have an overall higher GPA and more restricted grade range. See William D. Henderson, *The LSAT, Law School Exams, and Meritocracy: The Surprising and Undertheorized Role of Test-Taking Speed*, 82 TEX. L. REV. 975, 1005 tbl.7 (2004).

151. *Id.* at 1014 tbl.15.

152. The correlation coefficients were 0.347 for first-year required legal writing and 0.480 for in-class exams. *Id.* at 1012 tbl.13, 1014 tbl.14.

In sum, the LSAT does a decent, but not a great, job of predicting performance in law school, particularly when law school exams are timed essay exams. The type of test items that are particularly valid may have little bearing on our sense of the skills actually needed to practice law. A predictive validity model, however, does not claim to be valid as a direct measure of those skills themselves.

c. Speededness

The LSAT is commonly viewed as a speeded exam, with test preparation services offering coaching on how to improve performance by improving one's ability to do well in a limited time frame.¹⁵³ It is difficult to measure the speededness aspect of the LSAT directly because there is no penalty for guessing; hence, most students will mark an entry for each question. In the only rigorous study to date of a test comparable to the LSAT, the researchers concluded that the test score "was largely affected by the time limits."¹⁵⁴ More than half of the test takers (53%) found it necessary to engage in "rapid-guessing behavior" in order to complete the exam.¹⁵⁵

LSAC has kept track of the "speededness" of the various subcomponents of the exam, but as LSAC researchers recognized, their assessment is limited because there is no penalty for guessing.¹⁵⁶ The Analytical Reasoning section, for example, has been a mainstay of the exam since 1982. LSAC researchers concluded that it is not speeded because between 83-94% of all test takers completed this section of the test, and more than 80% of all test takers reached the last item of this section.¹⁵⁷ Because there is no penalty for guessing on the LSAT, however, one would expect that test takers would record an answer to every question even if they did not have time to answer the questions carefully. LSAC researchers also seemed inclined to conclude that a test item does not contain a speededness component even when the evidence appears to suggest otherwise.¹⁵⁸ For example, the "Reading Comprehension" section

153. HENDERSON, *supra* note 145, at 7.

154. *Id.* at 5 (reporting results of study by Schnipke and Scrans).

155. *Id.*

156. See REESE & COTTER, *supra* note 77, at 8:

It should be noted, however, that since LSAT scores are based on the number of questions a test taker answers correctly, many test takers realize that they are unwise to leave unanswered questions at the end of the test and should guess at items they do not have time to thoughtfully answer.

157. *Id.* at 9.

158. *Id.* at 9, 17.

“has appeared on more forms of the LSAT than any other item type.”¹⁵⁹ They concluded that “for the most part” this section was not speeded even though only 66% of the test takers completed this section of the exam during some administrations of the LSAT.¹⁶⁰

It is unclear whether LSAC abandoned certain test items because of strong evidence of speededness. The “Best Argument” item type, for example, was used from 1949-1951.¹⁶¹ It appeared as 30, 35, or 36 questions, and in some testing situations candidates were given only one minute per test item.¹⁶² In that version, only 52.4% of the test takers completed the section, and 80% reached only 22 of 30 items.¹⁶³ It was dropped after a few years of use.¹⁶⁴ An “Opposite” test item was added in 1950¹⁶⁵ and it was apparently designed to have a speededness component.¹⁶⁶ Eighty percent of the test takers reached 63 of 130 test items, and only 9.6% finished the section.¹⁶⁷ The “Data Interpretation” item type, however, was used from 1949-1976 despite strong evidence of speededness in some versions of this test item.¹⁶⁸ In its most speeded version, only 49.2% of test takers completed the section.¹⁶⁹

In a very thoughtful study¹⁷⁰ and article,¹⁷¹ William Henderson points out that examinations with a strong speededness component test for different abilities than untimed examinations.¹⁷² In other words, rank-order scores are likely to differ significantly depending on whether a test has a significant speededness component. “Within the field of psychometrics, test-taking speed and reasoning ability are viewed as distinct, separate abilities with little or no correlation.”¹⁷³ The speededness component of the LSAT, however, does not undermine its validity within the predictive validity model used by LSAC to validate the LSAT. Because speed is also an element of performance on in-

159. *Id.* at 17.

160. *Id.* at 17.

161. *Id.* at 3 tbls.1, 9.

162. *Id.* at 9.

163. *Id.*

164. *Id.* at 3 tbl.1.

165. *Id.* at 4 tbl.1.

166. *Id.* at 14.

167. *Id.* at 14.

168. *Id.* at 3 tbls.1, 10.

169. *Id.* at 10.

170. See HENDERSON, *supra* note 145.

171. See Henderson, *supra* note 150.

172. *Id.* at 979.

173. *Id.*

class timed law school exams, the validity of the LSAT is actually enhanced by having a speededness component.¹⁷⁴

The larger question, which is discussed in Part II, is whether it is valid for law school exams to have a significant speededness component. LSAC is simply trying to create an exam that will predict law school performance. So long as law school performance is determined, in part, by the ability to do well on timed testing instruments, LSAC will be compelled to include a speededness component in its testing instrument to improve its predictive validity coefficient. If LSAC, by contrast, defined the validity of its testing instrument with respect to untimed papers written during law school, it would have to redesign the test to achieve an acceptable validity coefficient.

One might argue that it doesn't really matter whether the LSAT has a speededness component because the content of the LSAT is largely arbitrary. Any combination of items, one might argue, is acceptable so long as those test items do a decent job of predicting who will do well in law school. That arbitrariness might be acceptable if it did not result in disparate impact against certain subgroups of law school candidates. In this Article, I argue that timed testing instruments are particularly difficult for students with certain learning disabilities as well as various psychological disabilities, such as ADHD or anxiety disorders. Some students with these disabilities request accommodations; others do not. If the test did not have a significant speededness component, then these requests for accommodation would become less necessary.

If the evidence were strong that speededness is an important component of success as a lawyer, then this disparate impact might be acceptable. But, as discussed below, there is no evidence that the type of speededness measured on the LSAT is important to the practice of law.

The speededness emphasis also causes another problem which will be discussed below—it causes some students to ask for extra time while taking the LSAT so that they can have an equal opportunity to demonstrate their skills and abilities. That accommodation, in turn, raises the question of whether their test score, with extra time, is a “valid” one. Does it predict first-year grades as well for that subgroup as for all test takers? If the LSAT were not an instrument with a significant speed component, many individuals with disabilities would not have to make the request for extra time, and the question of the validity of their scores under conditions of extra time would not arise.

174. *See id.*

ii. Students with Disabilities

LSAC takes the position that the LSAT is not a “valid” measuring instrument for individuals who take the exam under conditions of accommodation even though it is “valid” for all other test takers.¹⁷⁵ As discussed above, LSAC considers the 0.41 predictive validity coefficient sufficient to establish the exam’s validity for all test takers.¹⁷⁶ To understand its rationale for the conclusion that test scores are not “valid” for extra-time test takers, one must understand how accommodations take place and what data exists on students who receive extra time as an accommodation.

a. The Accommodation Application Process

Students with disabilities may request accommodations from LSDAS as part of the LSAT registration process.¹⁷⁷ LSAC began offering accommodations for the LSAT in 1982.¹⁷⁸ Although no formal studies have been done of the accommodation request process, data collected by LSAC suggests that the process has become more rigorous over the years. A technical report prepared in October 2001 reports the frequency of accommodation requests from 1993 to 1998.¹⁷⁹ Those data reflect that the number of requests for accommodations peaked in 1996-97, while the number of approved requests dropped from previous years. Subsequent to 1996, there has been a decline in the number of accommodation requests, which LSAC attributes to “administrative changes that led to more uniform guidelines for granting the approval of testing accommodations.”¹⁸⁰ In other words, the approval process became more rigorous and the number of accommodation requests declined.

Further, the amount of time extension provided also declined during this period. In June 1993, LSAC provided double time to all of the students who

175. THORNTON ET AL., PREDICTIVE VALIDITY, *supra* note 123, at 1.

176. See Selmi, *supra* note 116, at 1263-64; Brimer, *supra* note 122, at 43; see also *supra* note 120 and accompanying text.

177. Law Sch. Admission Council, Accommodated Testing, <http://www.lsac.org/LSAT/accommodated-testing.asp> (last visited Aug. 15, 2008) (describing the accommodation process for the SAT).

178. Laura F. Rothstein, *Bar Admissions and the Americans with Disabilities Act*, HOUS. LAW., Oct. 2004, at 34, 36, available at 32 OCT Hous. Law. 34 (Westlaw).

179. THORNTON ET AL., ACCOMMODATED TEST TAKER TRENDS, *supra* note 108, at 4 fig.1.

180. *Id.* at 4.

received time extensions.¹⁸¹ In October 1993, LSAC began to offer time-and-a-half rather than double time to some students.¹⁸² Starting with October 1996, LSAC began to offer time-and-a-half to more students than double time.¹⁸³ Thus, LSAC became more rigorous in its selection criteria and more stringent in the amount of extra time provided, beginning in 1996.

Today, a student who desires to take the LSAT under conditions of accommodation must undergo a rigorous application process. For the purposes of this discussion, I focus on students who seek extra time because of a cognitive impairment. The LSAC guidelines require that such students meet the following guidelines:

- The student must submit test results that are no more than three years old and reflect testing based on adult norms.
- The evaluator must be someone who has comprehensive training and direct experience in working with adult populations.
- The student must submit test scores from a comprehensive battery of exams and include a diagnostic interview, aptitude testing, achievement testing, information processing testing and personality testing; all actual test scores and subscores must be submitted, the report of assessment must include a specific diagnosis, and the report of assessment must recommend specific accommodations.
- The evaluator may *not* recommend that the LSAT be given as an untimed instrument.¹⁸⁴

In addition to those medical requirements, the applicant is required to complete a lengthy report about his accommodation history.¹⁸⁵ If a candidate, for example, ever had an Individualized Education Plan (IEP) under the Individuals with Disabilities Education Act (IDEA), then the candidate is expected to provide copies of past IEPs.¹⁸⁶ Because the IEP process is a yearly one, and IEP reports can be dozens of pages in length, a candidate could be required to provide over one-hundred pages of IEP documentation alone. The

181. *Id.* at 8 figs.5, 6.

182. *Id.*

183. *Id.*

184. Law Sch. Admission Council, Guidelines for Documentation of Cognitive Impairments 3, available at <http://lsac.org/pdfs/2008-2009/GuidelinesCognitive-2008.pdf> [hereinafter LSAC Guidelines]. Presumably, an expert also may *not* recommend that an individual receive a waiver from taking the LSAT altogether. Some blind students have asserted that, even under conditions of accommodation, their score is not predictive of their ability to perform well in law school because the logic games require the construction of charts, diagrams and graphs. Such skills are not necessary to performing well in law school but the lack of such skills could preclude admission to law school through a low LSAT score. See Naseem Stecker, *What's the Score: The LSAT and the Blind*, 80 MICH. B.J. 46 (2001).

185. LSAC Guidelines, *supra* note 184, at 1-5.

186. *See id.* at 1-2.

instructions also require the candidate to disclose whether she ever received accommodations for prior tests and what score she received on such tests.¹⁸⁷ If the candidate received accommodations in college, then the candidate must provide a letter from the Office of Disabilities on its official letterhead detailing all the accommodations the candidate ever received.¹⁸⁸

In order to request an accommodation, a candidate is also advised to provide this documentation well in advance of the registration deadline.¹⁸⁹ LSAC promises to reply to the candidate within two weeks of receiving the request and, if an accommodation is denied, allows a candidate to submit further documentation.¹⁹⁰ LSAC recommends that candidates start the accommodation request process six months in advance of the exam in case LSAC requires further documentation.¹⁹¹

These requirements make the accommodation process expensive. Students often receive a diagnosis of a disability at a young age and do not continue to have expensive diagnostic testing done to confirm the existence of the disability. And because they only recently reached the age of adulthood, most students would only have had testing done as a child. These rules require them to have a battery of tests completed after entering college. Few health insurance plans would cover such testing, and it could easily cost \$2,000 with the required extensive reporting requirements. If the students' parents had not retained their IEP reports, I do not know how a student would be able to collect such information.¹⁹²

The LSAC requirements also do not closely conform to the opinions of many educational experts about the diagnosis of a specific learning disability in reading. For example, LSAC requires that students provide results of a timed reading comprehension measure which has been normed on adults and recommends the Nelson-Denney Reading Skills Test (NDRT).¹⁹³ The NDRT, however, was designed to "quickly capture adult students' reading levels and screen for those who may benefit from remedial as well as accelerated

187. *Id.* at 2.

188. *Id.* at 1-2.

189. Law Sch. Admission Council, Accommodated Testing, *supra* note 177.

190. Law Sch. Admission Council, Accommodations Request Packet: General Information, <http://www.lsac.org/pdfs/2008-2009/AccommodationsFormGenInfo-2008.pdf> (last visited Aug. 15, 2008).

191. Law Sch. Admission Council, Accommodated Testing, *supra* note 177.

192. It seems ironic that LSAC requires adult test taking and adult norms yet also wants to see IEP reports from elementary school. If those data do not confirm the existence of a disability, then it is hard to know why it is relevant for this process.

193. LSAC Guidelines, *supra* note 184, at 2.

services. It is not validated for purposes of individual diagnosis.”¹⁹⁴ Researchers have found that the NDRT is not a good diagnostic tool; the Curriculum Based Measure (CBM) is a more reliable diagnostic instrument.¹⁹⁵ The LSAC requirements also do not appear to conform to current standards for administering and using other diagnostic instruments.¹⁹⁶

These extensive application procedures did not exist prior to 1997. Hence, it is not surprising that there was a dramatic drop-off in applications for accommodations after these practices were instituted.¹⁹⁷ The application process itself would seem to be a significant safeguard against fraud.

Despite the extensive documentation necessary to attain accommodations, LSAC takes the position that it cannot “validate” the accuracy of test results taken under conditions of extended time. Hence, their background material clearly states:

Candidates who seek additional test time on scored sections of the test should pay particular attention to the following:

- If you receive additional test time as an accommodation for your disability, LSAC will send a statement with your LSDAS Law School Reports advising that your score(s) should be interpreted with great sensitivity and flexibility.
- Scores earned with additional test time are reported individually and will not be averaged with standard-time scores or other nonstandard-time scores.
- Percentile ranks of nonstandard-time scores are not available and will not be reported.

194. Lewandowski et al., *supra* note 11, at 135.

195. *See id.* at 136.

196. *See, e.g.,* Earl S. Hishinuma, *Issues Related to WAIS-R Testing Modifications for Individuals with Learning Disabilities or Attention-Deficit/Hyperactivity Disorder*, 21 *LEARNING DISABILITY Q.* 228 (1998) (suggesting ways that the WAIS-R should be modified to accommodate students with learning disabilities).

197. LSAC reported the following data:

Year	1993-94	1994-95	1995-96	1996-97	1997-98
Requests for accommodation	No data available	1,489	1,560	1,645	1,425
Approved requests	1,195	1,348	1,255	1,087	822
Administered requests	970	1,113	951	798	630

THORNTON ET AL., *ACCOMMODATED TEST TAKER TRENDS*, *supra* note 108, at 4 fig.1.

- All information related to your request for accommodations will remain confidential unless you authorize its release.¹⁹⁸

In other words, candidates who take the LSAT with extra time, and have submitted extensive documentation to validate the appropriateness of extra time, receive a score that is sent to law schools. However, their LSAC report includes no percentile rank or index score.

The difference between test takers who are admitted to law school and those who are not may be particularly important in understanding the scores of accommodated test takers. Those who take the LSAT under conditions of accommodation may not be representative of all individuals with disabilities who seek to be admitted to law school. The very existence of the LSAT may have a significant screening effect on this population. This becomes clear by examining LSAC's own data. LSAC reported the following data for individuals whose requests for accommodation were approved.¹⁹⁹

	1993-1994	1994-1995	1995-1996	1996-1997	1997-1998
Approved requests	1,195	1,348	1,255	1,087	822
Administered requests	970 (81.2%)	1,113 (82.6%)	951 (75.8%)	798 (73.4%)	630 (76.7%)

As discussed above, these individuals submitted extensive paperwork well in advance of the test in order to have their accommodation requests approved. Nonetheless, in recent years, nearly one quarter of them do not bother to take the LSAT under conditions of accommodation (and presumably do not take the test at all). Why is that? LSAC offers no explanation for that trend, but there are some grounds for speculation. Quite possibly, such individuals conclude that, even with accommodations, they are unlikely to attain a score that reflects their ability to do well in law school. Hence, they decide not to take the LSAT at all. What is my basis for such speculation? My basis is that the drop-off in acceptance of accommodations (from 82% to 75%) coincides with LSAC typically offering candidates time-and-a-half rather than double time. When LSAC reports that such individuals had their

198. Law School Admission Council, Accommodated Test, <http://www.lsac.org/LSAT/accommodated-testing.asp> (last visited Aug. 15, 2008).

199. THORNTON ET AL., ACCOMMODATED TEST TAKER TRENDS, *supra* note 108, at 4 fig. 1.

requests “approved,” it is failing to note that, most likely, the full amount of accommodation they requested was not approved.

The difference between double time and time-and-a-half is substantial. The LSAT lasts three hours and twenty-five minutes. Double time results in six hours and fifty-five minutes. Time-and-a-half results in five hours and seven-and-a-half minutes. One can well imagine that some test takers might believe they cannot perform adequately unless they receive the additional one hour and forty-two-and-a-half minutes that is available to double-time test takers. In any event, we have no way of knowing the predictive validity of the scores of the nearly one-quarter of individuals who sought accommodations, were granted some accommodations, but chose not to take the exam. Not having such a large portion of students with disabilities in the predictive validity study would appear to skew the results.

b. Validity for Accommodated Test Takers

LSAC has reported the average LSAT score for non-accommodated test takers compared with test takers who received extra time as an accommodation. LSAC reported these data in graph form without reporting the precise score that corresponds to each datum point; nonetheless, it appears that the non-accommodated test takers, on average, scored 151 on the LSAT from January 1993 to February 1998.²⁰⁰ From January 1993 to February 1995, accommodated, extra-time test takers on average seemed to score one to two points higher than non-accommodated test takers.²⁰¹ Beginning in January 1995, when LSAC started to require additional documentation for accommodations and began offering less extra time, this difference diminished.²⁰² In fact, beginning in January 1996, this difference may have disappeared altogether.²⁰³ Unfortunately, LSAC grouped all of the data together to report that “Accommodated/Extra time test takers tended to have higher LSAT scores than the standard LSAT-taking population, on average” without noting the changes over time.²⁰⁴ But it does appear, at least for the time period in which double time was the standard extra-time accommodation, that accommodated test takers scored a point or two higher on the LSAT’s scaled score than non-accommodated test takers.

200. THORNTON ET AL., ACCOMMODATED TEST TAKER TRENDS, *supra* note 108, at 13 fig.11.

201. *Id.*

202. *Id.*

203. *Id.*

204. *Id.* at 12.

LSAC commissioned a study in 2001 to assess the “predictive validity” of accommodated LSAT scores.²⁰⁵ Despite evidence that the LSAT only has a predictive correlation coefficient of 0.41 for all test takers, LSAC presumed the general validity of the test and only assessed its differential validity—whether it was less valid for accommodated test takers.²⁰⁶ Did the LSAT scores of accommodated test takers predict their scores as well as those of non-accommodated test takers? And, in particular, were the scores predictively valid for those test takers who received extra time as an accommodation?

To answer this question, the researchers examined data from students entering law school from 1995 through 1999 because first-year grades would be available for those students.²⁰⁷ This discussion focuses on the data for students who received time accommodations.

The LSAC investigators focused their attention on the students who received accommodations and for whom complete data was available to assess predictive validity.²⁰⁸ Although the researchers did not report how many students received extra time accommodations during the years of their study, it is possible to estimate that the number was around 3,000.²⁰⁹ The study, however, was limited to those students who enrolled in law school and for whom an undergraduate GPA and first-year average from law school was available. That reduced the sample to 1,249 students.²¹⁰ No attempt was made to examine the LSAT scores of the 1,700 or so law school candidates who were not in the study. Not surprisingly, this group—which actually enrolled in law school and completed the first year—had a stronger profile than the typical LSAT exam taker. Their average LSAT score was 157.57 as compared with a 151 for the entire population.²¹¹ But as was found in a study of the entire population that takes the LSAT, the average scores of the students who received the accommodation of extra time were slightly higher than for the population as a whole—157.57 compared with 156.23.²¹²

205. See THORNTON ET AL., PREDICTIVE VALIDITY, *supra* note 123.

206. *Id.* at 1.

207. *Id.* at 3.

208. *Id.* at 1-3.

209. This estimate comes from the authors’ statement that the number of students receiving the accommodation of extra test time varied from 1,104 in 1994-95 to 610 in 1997-98. THORNTON ET AL., ACCOMMODATED TEST TAKER TRENDS, *supra* note 108, at 6.

210. THORNTON ET AL., PREDICTIVE VALIDITY, *supra* note 123, at 1-3.

211. *Id.* at 8 tbl.7.

212. *Id.*

For the sample of 1,249 students, the researchers then created a matched sample which consisted of students who did not receive accommodations but were otherwise matched for LSAT score, undergraduate GPA, gender, ethnicity, and year of entering class.²¹³ Because two-thirds of the extra-time test takers were male, the matched sample was also about two-thirds male. For some reason, though, the matched sample consisted of more Caucasians than the extra-time group. Whereas 78% of the candidates who received extra time were Caucasian, the matched sample was 96% Caucasian.²¹⁴ No explanation was offered for this discrepancy.

The researchers also concluded that the LSAT, when used alone or in combination with undergraduate GPA, over-predicted first-year grades in law school.²¹⁵ This conclusion is very important for students who receive extra-time accommodations on the LSAT because it justifies LSAC not providing an index score or percentile score for them in their LSAT report which is sent to law schools. If the purpose of the index is to project first-year grades, LSAC takes the methodological position that it cannot provide an accurate index for students who took the LSAT with extra time accommodations because the correlation between their index score and first-year grades is far weaker than it is for non-accommodated students.

There are two problems with this conclusion. First, it is premised on the assumption that the index score is a good predictor for non-accommodated students. In fact, it is not a strong predictor for any group of students, although it is less accurate for students who receive extra time accommodations.²¹⁶ But how much worse does it have to be in order for it not to be a “valid” predictor? For all students, the LSAT score correlation coefficient is 0.41; for students who receive extra-time accommodation, it is 0.34.²¹⁷ In other words, the LSAT accounts for about 16% of the variance in first-year grades for non-accommodated students and about 11.6% for students who receive extra-time accommodation.

That question leads to the second problem. LSAC also knows, through other validity studies that it has conducted, that LSAT scores also under-predict the first-year grades of female law students.²¹⁸ Yet, LSAC uses one

213. *Id.* at 6 tbl.6.

214. *Id.*

215. *Id.* at 9.

216. *Id.*

217. *Id.* at 8 tbl.7.

218. JENNIFER R. DUFFY & PETER J. PASHLEY, ANALYSIS OF DIFFERENTIAL PREDICTION OF LAW SCHOOL PERFORMANCE BY GENDER SUBGROUPS BASED ON 1999-2001 ENTERING LAW SCHOOL CLASSES 6 (Law Sch. Admission Council Technical Report 03-04, 2006), available at <http://www.lsacnet.org/>

percentile index for both men and women, suggesting that the scores are equally meaningful for both groups.²¹⁹ Further, as *U.S. News & World Report* has given increased weight to LSAT scores in its rankings, many law schools have placed greater emphasis on LSAT scores. The differential validity of the LSAT score, therefore, harms female law school candidates. Nonetheless, LSAC feels no need to delete LSAT percentiles for men out of the fear that their results over-predict their performance. Why not? Further, their studies confirm that, for students who take the LSAT more than once, the average of their LSAT test scores is a better predictor of first-year grades than either score standing alone.²²⁰ Yet, LSAC provides students with a predictive index based on *each* of their LSAT scores while knowing methodologically that *neither* of those results is the most accurate predictor. Again, one must wonder why LSAC allows individuals who have taken the test twice (and typically improved on the second test taking) to receive an index based on their higher test score even if that result over-predicts their first-year grades. Index scores are only withheld for students who receive extra-time accommodations based on a methodologically flawed study, as will be discussed below. Over-prediction is apparently acceptable for male candidates and for students who can afford to take the LSAT twice, but it is not acceptable for students with disabilities.

In any event, the conclusion that the LSAT over-predicts first-year grades for candidates who receive extra-time accommodations is not surprising because first-year exams are frequently time-pressured. The researchers did not control for whether students received extra-time accommodations during their first year of law school.²²¹ If they did not receive such accommodations, one would expect the LSAT score to over-predict first-year grades. This result is also not surprising because the researchers included data from the years in which the standard accommodation was double time, as well as years in which the standard accommodation was time-and-a-half. Beginning in October 1996, there was a sharp drop-off in the number of students who

Research/Analysis-of-Differential-Prediction-of-Law-School-Performance-by-Gender-Subgroups-Based-on-1999%E2%80%932001-Entering-Classes.pdf (“These results suggest that when LSAT alone is used as the predictor, the [First Year Average] mean for female students is more likely to be underpredicted and the [First Year Average] mean for male students is more likely to be overpredicted.”).

219. THORNTON ET AL., PREDICTIVE VALIDITY, *supra* note 123, at 7.

220. See SUSAN P. DALESSANDRO & LORE D. MCLEOD, THE VALIDITY OF LAW SCHOOL ADMISSION TEST SCORES FOR REPEATERS: A REPLICATION 10 (Law Sch. Admission Council Technical Report 98-05, 1998) (“[T]he most accurate use of the LSAT score is found by assigning the average LSAT score for the repeaters.”).

221. See generally THORNTON ET AL., PREDICTIVE VALIDITY, *supra* note 123.

received an accommodation of double time as compared with time-and-a-half.²²² Other data reveal that the average scores of accommodated test takers declined slightly after those changes were instituted. Whereas accommodated test takers had been scoring a point or two higher, on average, than non-accommodated test takers until February 1996, those differences virtually disappeared after June 1996.²²³ Thus, it would be helpful to know if the predictive validity of the LSAT score for accommodated test takers improved after February 1996.²²⁴

The more surprising result from the LSAC data is that grade-point average is *also* an unreliable predictor of first-year grades for students who receive accommodations on the LSAT.²²⁵ And this fact was true even though students who received extra-time accommodations on the LSAT, on average, had *lower* undergraduate grade-point averages than non-accommodated students (3.16 compared with 3.23).²²⁶ Even more surprisingly, undergraduate grade-point average over-predicted first-year grades both for students who received extra-time accommodations on the LSAT and for students who received accommodations other than extra time on the LSAT.²²⁷

222. THORNTON ET AL., ACCOMMODATED TEST TAKER TRENDS, *supra* note 108, at 13 fig.11. The numbers are reported on a graph so it is hard to determine the exact raw numbers that correspond with each point on the graph.

223. *Id.*

224. A recent study conducted for the College Board suggests that the LSAT is as valid a predictor of first-year grades for students who take the exam under extended time conditions as for other students. Researchers concluded that predictive validity scores of 0.48 for those who received no accommodation and scores of 0.35 for those who received accommodation were sufficiently close to make the SAT a valid predictor of first-year undergraduate grades for both groups. CARA CAHALAN ET AL., PREDICTIVE VALIDITY OF SAT I: REASONING TEST FOR TEST-TAKERS WITH LEARNING DISABILITIES AND EXTENDED TIME ACCOMMODATIONS, at 8-9 (College Board Research Report No. 2002-5), available at http://professionals.collegeboard.com/research/pdf/rr20025_11437.pdf ("Although the degree of correlation varies, the revised SAT is positively correlated with college achievement (as measured by first year grade point average) for students taking the test with extended time accommodations for a learning disability. SAT scores were fairly accurate predictors of FGPA for students with learning disabilities."). The College Board no longer "flags" the test scores of students who receive extended time; the 0.35 value was considered sufficient evidence of reliability as to not require flagging. The 0.34 value for accommodated test takers on the LSAT, as compared with the 0.41 value for non-accommodated test takers, is actually closer than the values deemed acceptable for the SAT not to be "flagged." A similar result with respect to the appropriateness of flagging accommodated test scores on the SAT was reached by a "blue ribbon" panel convened by the Northern District of California to study this issue. See generally NOEL GREGG ET AL., THE FLAGGING TEST SCORES OF INDIVIDUALS WITH DISABILITIES WHO ARE GRANTED THE ACCOMMODATION OF EXTENDED TIME: A REPORT OF THE MAJORITY OPINION OF THE BLUE RIBBON PANEL ON FLAGGING (2002), available at http://www.draregal.org/downloads/cases/breimhorst/majority_report.txt.

225. THORNTON ET AL., PREDICTIVE VALIDITY, *supra* note 123, at 9.

226. *Id.* at 8 tbl.7.

227. *Id.* at 9-10.

It is clear that this is correlation, not causation, data. Receiving accommodations on the LSAT does not *cause* a student to get grades that do not accurately predict law school performance. But remember the documentation that LSAC requires a student to complete in order to receive accommodations on the LSAT: the documentation requirements imply that students are more likely to attain accommodations on the LSAT if they had received accommodations during college for a documented disability. Thus, the undergraduate grade-point average was also likely to have been attained under conditions of accommodation which would presumably have the effect of raising it. When LSAC reports undergraduate grade-point average, however, it does not indicate that those grades were received under conditions of accommodation even though its own data reflects that those grades are as weak a predictor of first-year grades in law school as LSAT scores. For the group that receives accommodations other than time accommodations, the data indicates that the undergraduate grade-point average “provides the largest overprediction results.”²²⁸ It is interesting that LSAC does not emphasize this fact in any of its literature. It is careful to emphasize that the LSAT scores for this group are not accurate predictors of first-year grades, but it neglects to note that undergraduate grade-point average has its own set of problems as an accurate predictor. Ironically, the failure to provide a percentile to a student’s LSAT score may actually cause admissions offices to place undue weight on the undergraduate GPA of these students when, in fact, the LSAT is a better predictor of their first-year grades.

What is going on here? Why all this “over-prediction”? It is easy to blame LSAC for generating an exam that has too much time pressure and therefore depresses the scores of students with learning disabilities unless they are able to receive accommodations. And there is certainly evidence to support that hypothesis. When a student with a disability takes the LSAT twice—once without accommodations and once with accommodations—his score, on average, improves dramatically. Whereas the typical test taker improves by 2.7 points when retaking the exam, the improvement increases to 8.49 scaled score points when an individual changes from a non-accommodated to an accommodated exam format with extra time.²²⁹ In other words, a failure to receive an extra-time accommodation results, on average, in a decline in test score of five scaled points. For students at the middle of the testing range, a change from a 151 to a 156 is a jump from the 43.4

228. *Id.* at 9.

229. THORNTON ET AL., ACCOMMODATED TEST TAKER TRENDS, *supra* note 108, at 13.

percentile to the 64.2 percentile.²³⁰ Similarly, a jump from 146 to 151 corresponds to a jump from the 26.3 percentile to the 43.4 percentile.²³¹ For students in the middle of the pack, a small change in the number of right answers has an enormous effect on percentile ranking; the decision on whether to provide extra time has a correspondingly large effect on a student's LSAT score. The crude "on-off" nature of the accommodation decision, and an equally crude decision to give time-and-a-half or double time—without a precise scientific calculation as to how much time is truly warranted—leads to very imprecise LSAT test results that are used as if they are precise measures.

But an aspect of the over-prediction problem also lies with first-year teachers who overly emphasize time-pressured testing instruments. The fact that undergraduate GPA over-predicts first-year grades for students who receive accommodations on the LSAT suggests that these students were able to attain accommodations during college to better demonstrate their understanding of the material. Further, they were probably able to demonstrate their knowledge through a greater variety of testing instruments, many of which were probably not time-pressured. Law school, by contrast, especially in the first year, has a rigid uniformity that results in most students being tested with similar testing instruments. Hence, the students who do well during the first year of law school tend to be the students who do well on an LSAT-type testing instrument—a time-pressured exam.

As discussed above, the LSAT is a weaker predictor of grades when professors administer take-home exams or require papers. If the first-year curriculum were radically modified to emphasize take-home exams and papers, the overall predictive validity of the LSAT would likely decline, although its validity for students who receive accommodations might improve.

III. LAW SCHOOL EXAMS

A. *The Development of the Law School Exam*

The essay-style examination that is common in law schools today appears to be of long-standing existence—and critique. The *Columbia Law Review* published a series of articles on law school examinations in the mid-1920s.

230. See LSAC Test Form, *supra* note 97, at 38 (Conversion Table).

231. *Id.*

These articles focused on the subjective elements of essay exams and suggested that professors seek to move to more objective exam questions.

Professor Ben D. Wood wrote the first of several articles on *The Measurement of Law School Work* in 1924.²³² He was very critical of the subjectivity and inconsistency of the grading of law school essay exams. He observed that prior studies had demonstrated a correlation of 0.60 between grades in French and Trigonometry, yet he found correlations between various law school courses to be around 0.40.²³³ He also examined the predictability of class standing in the second year of law school based on performance during the first year. The reliability coefficient was 0.70, which he considered to be a poor measure of reliability.²³⁴ Correlation with grades earned as undergraduates at Columbia University or with Intelligence Test Scores was also low.²³⁵ Professor Wood did, however, locate one factor that seemed to improve the reliability of law school examinations. He found that the reliability of the testing instrument improved as the number of questions increased.²³⁶ No law school exam had more than eight questions, and reliability for an exam of that length was 0.71.²³⁷

Based on his investigation, Professor Wood concluded that professors should use some true-false questions rather than rely entirely on essay examinations.²³⁸ He found that the reliability of true-false tests was higher than the reliability of essay exams.²³⁹ Further, he found that the results from true-false tests correlated much more strongly with the results from intelligence tests than essay-style exams.²⁴⁰ But he also found that students needed to be given 200 true-false questions to answer in a two-hour period for the test to be highly reliable.²⁴¹ His study gave brief consideration to whether

232. Ben D. Wood, *The Measurement of Law School Work* (pt. 1), 24 COLUM. L. REV. 224 (1924) [hereinafter Wood (pt. 1)].

233. *Id.* at 243 tbl.6.

234. *Id.* at 244 (“The error of estimate is over 70% of what the error would be if the estimates were made on the basis of tossing a coin or some other pure chance basis.”).

235. The correlation with grades received at the undergraduate level was higher in the first year (0.47) and lower in the third year (0.30). *Id.* at 244. The correlation with Intelligence Test Scores ranged from 0.42 to 0.57. *Id.* at 245 tbl.7. He considered those correlations to be low although they are higher than the correlation today found between LSAT and first-year grades in law school. STILWELL ET AL., *supra* note 130, at 5-6.

236. Wood (pt. 1), *supra* note 230, at 248.

237. *Id.*

238. *Id.* at 247, 252.

239. *Id.* at 251 tbl.11.

240. *Id.* at 250.

241. *Id.* at 252.

“speededness” was affecting these test scores. He noted that less than a dozen students failed to answer all 200 questions and those students were typically students who also received a *D* or *F* on the essay part of the final examination.²⁴² He therefore suggested that their inability to complete the questions was reflective of their inability to master the material.

Professor Wood’s findings were very provocative because he was suggesting that more than half of the traditional law school exam be replaced with 200 true-false questions. He demonstrated that such an exam would correlate better with intelligence tests and undergraduate grades than traditional essay exams. Moreover, such exams would be more reliable in terms of providing more consistent grades from professor to professor.²⁴³ His mode of analysis, in some ways, is the reverse of the mode of analysis used in thinking about the LSAT. He was concerned that law school exams did not correlate strongly with intelligence instruments. Therefore, he suggested that the exams be modified to offer a better correlation with such testing instruments.²⁴⁴ He also encouraged law schools to administer intelligence tests to all incoming law students so that the reliability of law school exams could be measured in light of those testing instruments.²⁴⁵ He therefore presumed the validity of intelligence instruments and sought law school exams that replicated performance on such instruments.

But, of course, that line of reasoning raises the larger question of whether the intelligence instruments are valid, and, if so, whether we should expect law school grades to correlate with such testing instruments. If we are convinced that good lawyers are those who perform well on intelligence tests, we could dispense with law school grades and merely rank order students on the basis of such instruments. The fact that true-false test results correlate better with intelligence tests than essay exams does not tell us which instrument better measures who would be more successful at the practice of law. Professor Wood criticized law school essay exams because

[they judge] knowledge and ability from hastily written prose. . . . [T]he impromptu essay is, psychologically, hardly more a mirror of a student’s knowledge of and ability to think in the law of trusts than an impromptu speech would be. The case of the impromptu essay is aggravated by the slender samplings of materials and of performances which the time limits and its necessary brevity entail.²⁴⁶

242. *Id.*

243. *Id.* at 252.

244. *Id.* at 250-51.

245. *Id.* at 265.

246. *Id.* at 230 (footnote omitted).

Professor Wood's essay, however, made an important observation. He noted that essay exams are more reliable to the extent that they can cover more material.²⁴⁷ But the need to confine such tests to a few hours and restrict the amount of time that a professor will spend grading such exams apparently precludes extending the exam to deal with these problems. In the preface to Wood's essay, Dean Harlan Stone observed:

When I reflect that in the past dozen years I have devoted at least a solid year of working time, probably much more, to the painstaking reading of the thousands of examination books which have come to my desk in that period, I have reason to realize the sacrifice which law teachers are making to maintain the existing standards of measurement of law school work.²⁴⁸

With Dean Stone having already suggested that law professors spend too much time grading exams, it did not seem plausible for Professor Wood to propose a solution that would lengthen the exam process. Nonetheless, lengthening the standard three- to four-hour essay exam so that students respond to more questions could also certainly have increased the reliability of the testing instrument without changing its basic structure to a true-false exam.

Professor Wood's study also raised the larger question of how much consistency we should expect between aptitude tests, undergraduate grade-point average or scores on other law school exams, and the grade on a particular exam. Is a high correlation evidence of a better testing instrument? If the practice of law requires a variety of skills and abilities, then possibly we should not expect high correlation figures between these various measures of ability or learning. We require students to take a wide variety of courses presumably because we think these courses teach different content and skills that would be useful to lawyers. Hence, one could say that the measure of a well-rounded curriculum is that students obtain a variety of grades in different courses as they learn a variety of skills and content. If there is high inter-course reliability, we might wonder if the tests are simply measuring a

247. See *id.* at 229.

248. Harlan F. Stone, *Foreword to the Measurement of Law School Work* (pt. 1), 24 COLUM. L. REV. 221, 222 (1924). The desire to find ways of testing that save the time of law professors was also highlighted in the Foreword written by Dean Huger W. Jervy, introducing Professor Wood's second contribution to this study of the use of true-false questions. Huger W. Jervy, Foreword, *The Measurement of Law School Work* (pt. 2), 25 COLUM. L. REV. 316 (1925).

student's inherent ability rather than whether the student mastered the material taught in a specific course.

In his second installment of *The Measurement of Law School Work*,²⁴⁹ Professor Wood refined his suggestions somewhat. He concluded that true-false exams should be used in conjunction with essay exams, and that essay exams should provide students with a choice of essay topics:

By giving the student a choice of two problems out of four or five alternatives, we increase the chances that he really knows the facts necessary for the solution of his chosen problems. If the student knows the pertinent facts and fails to present a cogent and convincing solution, the teacher may with much greater assurance lay his failure to lack of reasoning ability. If, on the other hand, the student is compelled to write on three out of three questions, he may fail on one or even two, not for lack of reasoning ability, but because he chanced not to know the particular rule or rules pertinent to a particular feature of the problem. Reducing the number of questions on an Essay examination without giving the student a choice of problems will inevitably weaken the essay examination as a measure of reasoning ability unless the problems are so carefully selected by the Professor as to insure that every student who deserves to pass knows the facts and rules pertinent to a satisfactory solution. Otherwise, the Essay examination of two or three required problems is merely a more unreliable test of information than one of eight or ten problems.²⁵⁰

Professor Wood's recommendation about the essay exam was derived from his assumption that the entire testing instrument must be confined to three or four hours.²⁵¹ Because the true-false questions take two or so hours to answer, that leaves only about one-and-a-half hours for essay questions.²⁵² In that brief time period, it isn't possible to have students answer more than two questions. But Professor Wood had previously discovered that essay exams are more reliable if they contain more questions. Hence, his suggestion that professors use two hours of true-false questions negatively implicates how essay questions can be used. He concluded that the true-false questions can test breadth of knowledge, and the essay exam—by giving students an element of choice—can test reasoning ability. There is no need to use eight essay questions to test breadth of knowledge if the true-false questions have already tested that element.

In essence, Wood was suggesting that the true-false questions and the essay questions could serve different purposes. His data in support of that

249. Ben D. Wood, *The Measurement of Law School Work* (pt. 2), 25 COLUM. L. REV. 316 (1925) [hereinafter Wood (pt. 2)].

250. *Id.* at 319-20.

251. *Id.* at 321.

252. *Id.* at 321.

assertion, however, did not entirely fit his hypothesis. Wood demonstrated the reliability of the true-false questions by noting that they correlate much better with the scores on essay questions on other exams than do essay questions themselves.²⁵³ If the true-false questions test a different type of knowledge than essay tests, we might expect the true-false questions to correlate with each other but not with the results of the essay tests.²⁵⁴ Wood was also excited to report that the results of the true-false tests correlate better with the Thorndike Intelligence Test than do the essay tests.²⁵⁵ But it is not clear why that fact makes the true-false questions a particularly reliable measure of law school performance. It may simply reflect that there is more of an IQ component to performing well on true-false questions than essay questions. If the purpose of law school exams is to measure knowledge learned in law school, rather than pre-existing knowledge, one could argue that the essay exams are at least as valid a measure of performance in law school as the true-false questions.

Professor Michael Zander examined much of the evidence that Wood studied but came to different conclusions.²⁵⁶ Writing in 1968, he concluded that one disadvantage of the traditional essay exam is that it gives the candidate “little time to think.”²⁵⁷ He observed that “[o]ur form of examination system handicaps the student who is confused or overwrought by the strain and anxiety of exams. It gives inadequate or no credit for work done during the year. It does not test ability to do research even of the most superficial kind.”²⁵⁸ He concluded that a more appropriate testing instrument would be the “48 hour, 3,000 word answer” because it “can be an extremely valuable test of the capacity to think and write with some modicum of style in a lawyerlike manner.”²⁵⁹ He also recommended giving students “credit

253. *Id.* at 324-25.

254. In his third installment, Wood appears to have recognized the problem with the suggestion that true-false and essay exams measure different abilities. He noted the strong correlation between scores on true-false questions and scores on essay questions and stated: “The conclusion seems inescapable that the new-type examination is consistently both more reliable and more valid than the old-type, as to both knowledge and reasoning ability.” Ben D. Wood, *Measurement of Law School Work* (pt. 3), 27 COLUM. L. REV. 784, 790 (1927). It isn’t clear why Wood recommended the use of both the essay exam and the true-false exam if the true-false exam offers a more reliable indication of the same skills and abilities tested on the essay exam.

255. Wood (pt. 2), *supra* note 249, at 325.

256. Michael Zander, *Examinations in Law*, 10 J. SOC’Y PUB. TCHRS. L. 24 (1969).

257. *Id.* at 32.

258. *Id.*

259. *Id.* at 33.

towards the exams for work done during the course of the year.”²⁶⁰ Finally, he recommended that examinations break down the traditional barriers between subjects by requiring “students to take at least one paper which would cover a variety of fields and would test their diagnostic skills.”²⁶¹ Although these suggestions were made about forty years ago, few of them, as we will see, have been implemented in law schools today.

B. Today's Exams

i. Literature on Law School Exams

Professor Kissam argues that the grading of law school examinations has evolved from an Aristotelian model to an objective model.²⁶² Under the Aristotelian model, exams were graded “under a general or holistic approach that gave considerable emphasis to a professor’s practical judgment about the professional promise indicated by different student answers” and were generally accorded an A, B, or C, although D and F grades were also available.²⁶³ He describes this model as Aristotelian “because of its capacity to take account of the skills of interpretation, conventional and creative imagination, practical reason, and practical judgment, all of which are associated with Aristotle’s philosophy of ethical or normative decisionmaking.”²⁶⁴

With the modern pressures to rank order students on the basis of their examination answers, law school grading has gone towards a more objective model which Kissam describes as “piecemeal and fragmented,” because the professor searches for “the many specific elements to a ‘right answer’” and reads the exam “in a negative state of mind in order to produce the many quantitative distinctions that are inherent in objective grading practices.”²⁶⁵ He argues that this objective methodology rewards “a student’s abilities to perform the basic functions of issue spotting, rule specification, and rule application quickly and productively.”²⁶⁶ He criticizes these quantitative methods as giving “little if any consideration to the broader, more practical,

260. *Id.*

261. *Id.* at 34.

262. See Philip C. Kissam, *Law School Examinations*, 42 VAND. L. REV. 433 (1989).

263. *Id.* at 445.

264. *Id.* at 446.

265. *Id.* at 446-47.

266. *Id.* at 447.

and professionally oriented skills that can be recognized by the Aristotelian method.”²⁶⁷

The need to move to a rank-order system of grading has caused law professors to turn to time-pressured examination instruments. Such testing instruments can easily yield a wide range of student answers allowing a professor to comply with a curving requirement. But such examinations only test a limited range of abilities. “The principle of speed rewards students for answers that merely identify a maximum number of issues and specify precisely many different rules. Thi[s] principle tends to de-emphasize, discourage, and penalize student writing that involves coherence, depth, contextual richness, and imagination.”²⁶⁸

Various justifications are offered for the timed nature of law school exams and the “speededness” that they frequently require. When I gave a faculty workshop on the topic, these are some of the justifications that I heard:

- It reflects the real-world time pressures of being a lawyer;
- Clients pay lawyers by the hour, so “quick” lawyers are more valuable lawyers; and
- A timed testing instrument forces students to synthesize the material in advance of the exam and keep up with the material, the combination of which enhances learning.

Clinical professors, however, dispute that law school exams reflect the way time pressures operate in practice. For example, Professor Janet Motley, who is a clinical professor, observes that “[t]here is rarely an occasion in practice when [speed] is required.”²⁶⁹ Further, she argues that asking students “to speedily answer a complex legal problem” implies that that is the way they should practice law.²⁷⁰ “The enticement to take on an overly burdensome caseload is unchecked by regard for professional excellence when hasty problem-solving is the training norm.”²⁷¹ Professors, themselves, she observes, would never work under the kind of time constraints that they impose on students during the exam process.²⁷²

Admittedly, there are times in practice when it is advantageous to be quick. During depositions, or oral arguments, one often has to think of a

267. *Id.*

268. *Id.* at 453.

269. Janet Motley, *A Foolish Consistency: The Law School Exam*, 10 NOVA L.J. 723, 736 (1986).

270. *Id.*

271. *Id.*

272. *Id.* She specifically asks: “[H]ow many law professors spend merely the one to three hours allocated for an essay exam in drafting their model answers?” *Id.*

sequence of questions or arguments fairly quickly. But those are examples of verbal quickness, not quickness in writing. Further, many factors other than pure speed are involved in those contexts. Because of the verbal context, one may be reading interpersonal cues to assess what direction is appropriate. Additionally, one might need to be quick when making objections in the courtroom. That quickness requires certain listening skills that are not tested on law school exams. Law school exams best reflect memos that one might write to another lawyer or briefs that one might write to a judge. It's hard to think of too many circumstances where one would want to complete those documents in a few hours rather than devote at least eight hours to them. Even if one *could* complete a draft in a few hours, one would most likely want to turn in a more polished product. We teach students the importance of care and precision in law and then test them on their ability to turn in very unpolished drafts.

The better case for timed exams is that they force students to synthesize the material in advance of the exam. It is probably easier to procrastinate with respect to studying material if one knows that the exam will be a forty-hour take-home exam. But would a student be more likely to procrastinate if the exam were eight hours rather than four? Or what about five or six hours instead of three? Is the time limit merely administratively convenient rather than pedagogically appropriate?

Even if one accepts the assertion that a shorter exam causes students to synthesize the material in advance, one might ask whether other processes—other than three- to four-hour essay exams—could produce the same result. One underlying assumption with respect to law school grading is that one exam at the end of the semester is an appropriate way to construct nearly the entire grading profile. That choice of measuring instruments puts enormous pressure on students irrespective of whether it is four hours or four days in length. In college, students were accustomed to quizzes and mid-year exams, along with finals. Yet law school has always emphasized the single exam at the end of the year. A major justification for this examination process is efficiency. Unlike undergraduate professors who teach large courses, law professors do not get teaching assistants who assist in grading exams. Given the hours that it takes to grade essay exams, it is simply not realistic to expect law professors to engage in that exercise more than once during the semester. The lack of evaluation during the semester makes it more important for the final exam to be in a format that encourages students to be prepared throughout the semester. Increased use of multiple-choice competency quizzes or reading checks might be able to solve that problem without the use of time-pressured final exams. Or maybe we could even hire teaching

assistants to help us grade quizzes that involve some writing. It is even possible that law professors *could* read short essays throughout the semester and give the students some feedback, in advance of the exam, to encourage them to synthesize the material.²⁷³

The durability of the three- to four-hour essay exam as the primary determinant of one's grade is rather surprising given the changes that have occurred in legal education in the last twenty years. We have changed the first-year curriculum many times, added many clinical courses, started to teach alternative dispute resolution, and admitted a much more diverse class of students. Despite all of these changes, the law school exam has remained relatively static. It is time to re-examine our devotion to this testing instrument.

ii. A Modest Survey and Empirical Study of Exams

Today's law students are accustomed to a particular type of examination, especially in the first year of law school. The exam is typically a three- to four-hour exam primarily consisting of essays but possibly including some multiple choice or short objective questions. In some cases, the student may take a longer version of this exam for anywhere between six and twenty-four hours. Irrespective of the length of the examination period, the exam is likely to test students on their ability to spot issues, state doctrine, discuss the "gray" areas of the law, evaluate the strength of arguments, and see both sides of an argument. Sometimes, students are also asked to make policy arguments or assess what is the "best" argument.

I conducted a brief survey of one faculty's grading practices to assess the variety of exams given in that law school. For the purpose of this survey, I only considered courses in which the professor assessed most, if not all, of the grade through an examination. I used the file that the law school maintained for the American Bar Association Accreditation Committee on all exams administered in the law. During the fall 2006 semester, there were forty-seven

273. One option I have tried is to make practice questions available on The West Education Network (TWEN), located at <http://lawschool.westlaw.com>, in advance of the final. Students are encouraged to answer the questions anonymously and send me their answers to evaluate. I limit the word count on these practice questions to 1,000 words and find that it takes me no more than ten minutes to read a student response and offer some constructive suggestions. The availability of this opportunity encourages students to synthesize the material in advance of the final and also gives them a way to get quick feedback on how they approach exams. I usually can provide them with an assessment on the same day that I get the sample answer. Less than half of the class takes advantage of this opportunity to get feedback. Because the exam answers are submitted anonymously, I have no way of knowing if those students perform better on the final.

courses offered in which professors gave final examinations. Every exam administered to first-year students was an in-class examination. The exams varied in length from three hours to three hours and forty-five minutes. First-year students were divided into three sections and had final examinations in three courses. One of these classes had a smaller enrollment than the other two. Of the exams that students took during their first semester of law school, one or two contained some multiple choice or true-false questions; all exams contained essay questions. With respect to essays, most exams contained three essay questions although one exam contained as many as eleven essay questions.²⁷⁴

In the first semester, the use of objective questions in upper-class courses was comparable to their use in the first-year classes. Twelve of the thirty-two (37.5%) upper-class courses (with exams) included objective questions. The number of objective questions was also lower than in the first year with the range being from six to fifty. In the first-year courses, the range was from fifteen to two hundred. Take-home exams were also more prevalent among upper-class courses with nine of the thirty-two (28.1%) upper-class courses offering exams ranging from seven hours to the entire length of the exam period. The enrollment in classes with take-home exams tended to be smaller than other classes. No class with enrollment over forty offered a take-home exam. Because take-home exams were limited to comparatively small enrollment classes, most students would take courses with timed in-class exams.²⁷⁵

The pattern for the second semester was similar. The faculty offered forty courses with exams. The smaller number of courses with exams can be explained, in part, by the fact that all first-year students were required to take Legal Writing during the spring semester. Hence, there were more writing courses offered in the spring semester than in the fall semester. Of those forty courses with exams, twelve were first-year courses. (Each first-year student had four exams during the second semester.) Nine of the first-year courses were graded with a timed in-class exam, one had a twenty-eight hour exam, and two had an eight-hour exam. Only two of the twelve (16.7%) courses had

274. It is hard to compare or describe the essay questions because some professors asked single questions with many parts; other professors asked a series of short questions. Although it would be interesting to compare grades based on the type of essay questions administered, I have not done so for the purpose of this Article because of the difficulty of labeling the types of essays questions used.

275. This difference in enrollment between courses with in-class and take-home exams makes it difficult to compare the grades from these kinds of measuring instruments. Although no systematic data are kept on this subject, it is common knowledge that the grades in small-enrollment courses tend to be higher than the grades in classes with higher enrollment.

any multiple-choice questions; the other ten were based entirely on essays. The examination procedures varied considerably by section. One section had two take-home exams, one section had one take-home exam, and one section had no take-home exams. The section with two take-home exams also had no examinations that used objective questions (multiple choice or true-false) in the second semester.

The following chart reflects the variety of exams used for first-year students throughout the year:

	First Semester: use of objective questions	First Semester: use of take home exams	Second Semester: use of objective questions	Second Semester: use of take home exams
Section One	One small section had objective questions; otherwise, no objective questions	No take- home exams	Objective questions in one class	No take- home exams
Section Two	Two classes used objective questions	No take- home exams	No objective questions	Two take- home exams
Section Three	No objective questions	No take- home exams	Objective questions in one class	One take- home exam

Overall, there is a nice balance throughout the year with respect to the use of objective questions. Each student had one or two classes that used some objective questions on the final examination. Less consistency existed with respect to the use of take-home exams. Students were exposed to zero, one, or two courses with take-home exams. If student performance varies depending on the type of examination instrument, it would seem to be particularly important to offer consistency in the first year since first-year grades are a factor in the selection of law review members and also play an important role in employment opportunities. Although many students score consistently irrespective of the type of examination instrument, the type of

examination instrument is important for some students, as will be discussed below.

With respect to the upper-class curriculum, twenty-eight courses were offered with exams. Nine of twenty-eight (32.1%) courses offered some objective questions on the final, which was consistent with the figure for the fall (37.5%). Take-home exams were less prevalent in the spring than in the fall. Only five of twenty-eight (17.8%) offered take-home exams in the spring compared with 28.1% in the fall. Except for one course with forty-two students, all of the classes with take-home exams had enrollment of under forty students. Hence, the pattern of take-home exams being more prevalent in small enrollment courses continued.

The following chart summarizes the general distribution of take-home and in-class exams at the law school under study:

Distribution of Take-Home and In-Class Exams for Courses with Exams

	Fall 2006	Spring 2007
Take-Home Exams	9 (19%)	8 (20%)
In-Class Exams	38 (81%)	32 (80%)

As the table suggests, students at this law school are graded overwhelmingly on the basis of in-class timed exams for courses in which exams are given. The table does not include clinical courses, seminars, or other courses evaluated primarily by written work. Nonetheless, the table probably underestimates the prevalence of timed in-class exams for law students because it does not control for the size of the class. For classes with more than forty students, in-class exams appear to be nearly the universal method of examination after the first year. The class rank is determined predominantly by these in-class timed instruments.

It is difficult to estimate the time-pressured dimensions of these exams based merely on a description of their format, but it was apparent to me in reviewing the exams that some would be very time pressured. The most time-pressured exams were longer than ten pages, had many subparts to their questions, and offered the students no more than three hours or three-and-a-half hours to complete them. Many professors offered what they described as "extra" time beyond the time specified in a question. That extra time ranged from five to forty minutes, but it was hard to see how that extra time detracted

from the time-pressured aspects, especially for professors who insisted that students not begin writing until the “extra time” had expired. A review of the exams also suggests that the time pressure varies enormously on in-class exams. One professor gave the students four hours to answer two relatively short questions. Another professor gave the students three hours to answer two relatively short questions. By contrast, one professor asked students to answer thirteen questions in three-and-a-half hours. In another course, the three-hour exam was thirty-one pages in length. The take-home exams were probably less time pressured than the in-class exams, but the time pressures among the in-class exams did vary.

None of this variation in exams really matters unless the variation affects the resulting rank order among students. Other data that I have collected suggests this variation does affect the rank ordering of students.

To reach that conclusion, I examined test scores collected by a professor at a law school. She had her students for an entire year and gave them two examinations—one at the end of the fall semester and one at the end of the spring semester. The students were told that each exam was 50% of their grade. For various reasons, a few students each year did not take both exams and they were not included in this study. Enrollment in the classes was about sixty students.

I examined four years of data. In year one of the study, the professor gave what she self-identifies as a “very speeded” exam in the fall semester. She then gave a take-home exam in the spring semester. In year two, she gave an in-class exam in the fall semester, but she adjusted the exam to make it less speeded than in the previous year. She again gave a take-home exam in the spring semester. In year three, she gave an in-class exam in both semesters; these exams were similar to the in-class exam she gave in year two. In year four, she gave an in-class exam in the fall semester and a take-home in the spring semester. Once again, her in-class exam was the less speeded version that she began to use in year two.

I calculated the correlation coefficient in each year between the student scores on the fall semester exam with the scores on the spring semester exam. As predicted by the educational psychology literature, the correlation coefficient was highest when she used comparable testing instruments in the fall and spring (year three) and lowest when she used a highly speeded exam in the fall and a less speeded exam in the spring (year one).

Correlation Coefficients Between Fall and Spring Exams

	Types of Exams Compared	Pearson Coefficient	Correlation
Year One	Very speeded in-class exam and take-home exam	0.467	(22% variance)
Year Two	Less speeded in-class exam and take-home exam	0.626	(39% variance)
Year Three	Less speeded in-class exam and less-speeded in-class exam	0.753	(57% variance)
Year Four	Less speeded in-class exam and take-home exam	0.638	(41% variance)

These data offer two key insights. First, they support the hypothesis that the type of examination instrument used does affect the rank order. The lower correlation coefficient when two different types of exams are used suggests that different kinds of exams—specifically, timed and untimed instruments—do test for different skills and abilities. Further, these data suggest that professors can modify in-class instruments to make them less timed and more like a take-home exam. Finally, we should remember that these data are for *all* students. There are too few students in any class who are identified as having learning disabilities for me to consider their test scores separately. For students with learning disabilities who are not receiving extra-time accommodations, one would expect the correlation between in-class and take-home instruments to be even lower than those reported above for all students. And for students with learning disabilities who do receive extra time, there is the difficult question of how much extra time is appropriate. If interviewed, this professor might have indicated that her in-class exam in year one had a significant time pressure for students without disabilities but contained little time pressure in year three. Would it be possible to individualize the allocation of extra time so that all students are taking the exam under comparable conditions? If time-pressured in-class exams were not the overwhelmingly common type of exam administered in law school, obtaining the correct answer to that question would not be so important.

One last way to view the data of this professor, which also gives insight into the meaning of the LSAT's low correlation coefficient, is to compare the ranking of students in year one of this data. In year one, the correlation coefficient which seeks to predict the score on the spring semester exam from the score on the fall semester exam was similar to the correlation coefficient which seeks to predict the grades during the first year of law school from the

LSAT score. The LSAT correlation coefficient is 0.41; the correlation coefficient for this professor's students was 0.47 (a modest improvement).

In the table below, I report how the ranking of students would change depending on which exam was used for ranking purposes. I have deliberately only listed those students for whom the ranking would change considerably. My reporting below includes about one-fourth of the fifty-three students for whom I have complete data.

Student Percentile Ranking Based on Take-Home or In-Class Exam

Student Number	Take-Home Exam Ranking	In-Class Exam Ranking
1	66%	4% ²⁷⁶
2	15%	83%
3	19%	36%
4	17%	74%
5	96%	55%
6	2%	23%
7	13%	62%
8	77%	98%
9	36%	4%
10	92%	43%
11	60%	98%
12	55%	96%
13	2%	19%

276. In other words, 66% of the students performed better than this student on the take-home exam but only 4% of students performed better than this student on an in-class exam.

For some students, the change in performance was quite dramatic. The significant improvement on the untimed instrument compared with the timed instrument for students 2, 4, 6, 7, 8, 11 and 12 would cause me to wonder if those students might have an undiagnosed learning disability. Students 2 and 4, in particular, had wide swings in performance—more than two standard deviations. Of course, there are always some students who take a while to “catch on” during law school, and those results deserve further investigation to see if they are otherwise replicated in these students’ performance. Did these students generally have improvements from the fall to spring semester, or was this improvement isolated to a class in which the examination instrument changed from in-class exam to take-home exam? I would recommend that associate deans monitor student performance on the basis of the type of exam to see if testing for a learning disability is warranted.²⁷⁷ Unfortunately, there is so little variation in examination instruments at most law schools that such comparisons are not possible. If the performance of students 2 and 4 on this professor’s in-class timed instrument is consistent with their general performance on in-class timed instruments, we can see how the type of examination is very much affecting their employment prospects upon graduation. Whether they are in the top quartile or bottom quartile of their class may depend on the examination instruments used to evaluate their performance.

Returning to the subject of the LSAT, let us again consider students 2, 4, 6, 7, 11, and 12. How might their performance on the LSAT have improved had it been a less time-pressured instrument? Might they have been able to attend a more prestigious law school? Should some of them have requested extra time on the LSAT but were unaware of a learning disability?

IV. CONCLUSION

The standard response to the observation that some students’ disabilities cause them to have slower processing speeds is to suggest that these students be given extra time to take examinations. On exams in which the standard amount of time does not contain a “speededness” component for nondisabled test takers, that solution may be the appropriate one. The law school setting, however, poses special challenges because exams are graded on a rank-order

277. On two occasions, based on such data for individual students, I have suggested testing for a learning disability. In both cases, the students easily met the criteria for a learning disability upon being tested. Both students came from less privileged backgrounds where testing would have been unlikely.

basis and typically have a speededness component. Similarly, the LSAT is graded on a scaled score and has a speededness component. If students with disabilities are graded on the basis of an essentially “unspeeded” instrument, then we might expect their test scores to be comparatively higher than those of students who learned the material equally well but took the exam under speeded conditions. It is possible to adjust the allocation of extra time to simulate the same degree of speededness for all test takers, and LSAC may have achieved that result in recent years on the LSAT. But such precision is challenging to accomplish.

Other solutions are also possible, although none is free of problems. The standard amount of time that is allocated for all test takers could be lengthened to lessen the speededness component. Then, when students with disabilities are given extra time, both sets of students are taking an essentially unspeeded exam. Three problems exist with this solution. First, it is not clear that extending the time of law school exams from four hours to eight hours, or even twenty-eight hours, truly eliminates the speeded element for nondisabled test takers. Second, there are administrative challenges in allowing more than eight hours per exam because the exam period might have to extend over several weeks. Self-scheduling of exams could solve that problem, but that solution raises security concerns. Finally, some students might not study as effectively throughout the semester if they think they had a lot of time to complete the exam.

Another solution would be for law schools (and LSAC) to stop grading exams on a rank-order basis to lessen the challenges of making comparisons between law students. Employers and others, however, might object to law schools changing their grading system. And, once again, students might study less effectively if grades on exams were considered less important.

Finally, law schools could change their examination system overall so that it gives less weight to a timed end-of-semester exam. Throughout the semester, students could take untimed, multiple-choice competency exams and write short papers, and be graded more significantly on the basis of class participation or oral projects. One of the key advantages of the current system is convenience for law faculty—grading one exam at the end of the semester is far easier than grading numerous instruments throughout the semester. That convenience, however, may not serve a pedagogically sound method of grading. The end-of-semester exam would not have to be as comprehensive if other testing mechanisms are used throughout the semester. Nonetheless, with first-year class sizes often hovering around seventy to eighty students, it would probably be difficult to persuade law faculty to change their grading systems.

Although no solution is perfect, my research does strongly suggest that there is a fairness problem with all faculty members giving comparable speeded exams during law school, especially in the first year, when grades typically also serve the purpose of determining law review standing. Processing speed clearly affects the performance of some students, and not all of those students receive accommodations during law school. Thus, at a minimum, it seems important to grade students within a section on the basis of more than one type of examination instrument. Ideally, half of their exams would be take-home exams and half would be in-class exams.

Adopting a variety of law school exam formats is not simply fair to students with disabilities; it is good exam procedure for all law students. The practice of law involves a wide variety of skills. If all law school exams have a strong speededness component, we are not measuring that variety of skills—we are testing speededness over and over again.

Because we allow law faculty substantial discretion in the construction of exams, it is difficult to implement the recommendations made in this Article. Associate deans and others involved in curriculum matters, however, can examine their law school's testing practices as a whole and see whether they have struck the right balance, particularly in the first year. Subtle changes could produce better results for everyone by allocating faculty to first-year classes, in part, based on their testing practices. This Article suggests that law schools have an obligation to examine their testing practices and to see how they could be improved to attain greater fairness in the allocation of grades. We teach our students the importance of the principle of fairness; we can set a good example by implementing that principle ourselves.