

Test Validity: Faster is Not Necessarily Better

Ruth Colker*

This Article argues that we should change the default rule for standardized testing: Test developers should not be allowed to implement speeded exams that adversely impact individuals with disabilities unless test developers can validate the time limits. Applying principles developed under Title VII of the Civil Rights Act of 1964 to the Americans with Disabilities Act, test developers should need to demonstrate that a “speeded,” standardized exam is a valid predictor of the desired skills and abilities, and that no less-impactful alternative, such as a “non-speeded exam,” is available to measure those skills and abilities. The shifting of the default rule to non-speeded exams would mark the implementation of a new, universal design principle that would make standardized testing more equitable for a range of people, including racial minorities, women, people with low socio-economic status, older applicants, and individuals with disabilities. Testing entities should devise non-speeded exams that can validly measure the skills and abilities of the entire applicant pool, rather than continue to place the burden on people with disabilities to meet onerous and expensive standards to request extended time. This solution is novel, simple, and fair.

* Distinguished University Professor and Heck-Faust Memorial Chair in Constitutional Law, Moritz College of Law, The Ohio State University. I would like to thank Moritz Law Librarian Stephanie Ziegler for her assistance in collecting some of the sources cited in this Article, Ric Simmons, Jasmine Harris, Dan Tokaji, Dakota Rudesill, Sasha Samberg-Champion, Nancy Mathers, and David Levine for their helpful feedback on an earlier draft, and Professor Guy Rub for administering a survey to our students about their experiences taking the LSAT. I would also like to thank the faculty at University of California, Davis, and the Moritz College of Law for their helpful comments at a faculty workshop. Finally, I would like to thank the staff at the Law School Admissions Council who met with me on June 20, 2018, to discuss many of the ideas presented in this Article and are currently considering a grant proposal to assess the validity of the speeded aspect of the LSAT.

680

SETON HALL LAW REVIEW

[Vol. 49:679]

I. INTRODUCTION	681
II. SPEED	689
A. Speeded Exams.....	689
1. Introduction.....	689
2. Maximum Potential Thesis	692
3. Differential Boost Thesis	693
4. Flaws and Difficulties.....	695
5. Response to Flaws and Difficulties	698
6. A Promising Model.....	702
B. LSAT as a Speeded Exam	703
1. Introduction.....	703
2. Current LSAT Discussion.....	706
III. TITLE VII PRINCIPLES	713
A. Employment Context.....	713
B. Outside the Employment Context.....	718
IV. ADA PRINCIPLES	719
V. A TITLE VII/ADA SYNTHESIS.....	723
A. Importing Title VII Validity Principles into Section 12189	723
B. Application of the Test Validity Approach to the LSAT	727
1. Existing Validity Data	727
2. Arguments Against Validity of LSAT.....	729
3. Response to Arguments in Favor of a Speeded Exam.....	731
VI. CONCLUSION.....	735

I. INTRODUCTION

We are at a moment of crisis in legal education. The courts are possibly on the verge of banning race-based affirmative action.¹ Law schools offer merit-based grants to prospective students to raise law schools' median Law School Admission Test (LSAT) scores, exacerbating the lack of socioeconomic diversity in law schools.² Law schools are beginning to experiment with selecting applicants on the basis of Graduate Record Examinations (GRE) scores,³ even though this math-heavy exam is weighted in favor of male applicants.⁴ And the news media has hammered the allocation of extended time to students with disabilities⁵ at a time when a

¹ In a four-to-four decision affirming the Court of Appeals, the United States Supreme Court upheld the University of Texas's limited, race-based affirmative action plan for admissions in an opinion authored by Justice Anthony Kennedy. *Fisher v. Univ. of Tex.*, 136 S. Ct. 2198 (2016). With Justice Antonin Scalia's replacement on the Court by Justice Neil Gorsuch, and Justice Kennedy's replacement by Justice Brett Kavanaugh, such plans are in legal jeopardy. While Justice Gorsuch has not specifically ruled on affirmative action, one would expect his conservative legal philosophy to align more closely with Justice Scalia's than with Justice Kennedy's. See Mark Sherman, *Gorsuch Establishes Conservative Cred in 1st Year on Court*, ST. AUGUSTINE REC. (Nov. 26, 2017), <https://www.staugustine.com/2017-11-26/gorsuch-establishes-conservative-cred-1st-year-court>. Similarly, one would expect Justice Kavanaugh to side with the conservatives on the Court. See Adam Liptak, *How Brett Kavanaugh Would Transform the Supreme Court*, N.Y. TIMES (Sept. 2, 2018), <https://www.nytimes.com/2018/09/02/us/politics/judge-kavanaugh-supreme-court-justices.html>. Students for Fair Admissions is aggressively pursuing litigation against various universities to argue that their affirmative action programs are illegal or unconstitutional. See *Update: Harvard's Education in Discrimination*, STUDENTS FOR FAIR ADMISSIONS, <https://studentsforfairadmissions.org/harvards-education-in-discrimination/> (last visited Nov. 21, 2018) (describing its lawsuit against Harvard University).

² Merit-based financial aid causes financial aid dollars to flow to high-income students, moving scholarship dollars from the poor to the wealthy and middle class. See Meredith Kolodner, *States Moving College Scholarship Money away from the Poor, to the Wealthy and Middle Class*, HECHINGER REP. (June 22, 2015), <https://hechingerreport.org/states-moving-college-scholarship-money-away-from-the-poor-to-the-wealthy-and-middle-class/> (reporting that "[t]welve states plus Washington D.C. now spend more on merit-based aid than need-based aid"). For a discussion of the use of merit-based financial aid among law schools, see Farran Powell, *Law Schools Shell Out Deep Tuition Discounts to Students*, U.S. NEWS & WORLD REP. (Nov. 20, 2017), <https://www.usnews.com/education/best-graduate-schools/top-law-schools/paying/articles/2017-11-20/law-schools-shell-out-deep-tuition-discounts-to-students> (reporting that "[s]tudents with high LSAT scores could see as much as \$100,000 in tuition discounts").

³ See *Taking the GRE General Test for Law School*, EDUC. TESTING SERV., https://www.ets.org/gre/revised_general/about/law/ (last visited Feb. 13, 2019) (listing schools that accept GRE scores for law school admissions).

⁴ See *infra* Part II.B.2.

⁵ See, e.g., Douglas Belkin, *Colleges Bend the Rules for More Students, Give Them Extra Help*, WALL ST. J. (May 24, 2018), <https://www.wsj.com/articles/colleges-bend-the-rules-for-more-students-give-them-extra-help-1527154200>; Doree Lewak, *Rich Parents Are Using Doctor's Notes to Help Kids Cheat the SATs*, N.Y. POST (May 2, 2018), <https://nypost.com/2018/05/02/rich-parents-are-using-doctors-notes-to-help-kids-cheat-the-sats/>; Ari Trachtenberg, *Extra Time on an Exam: Suitable Accommodation or Legal*

presidential candidate openly mocked a reporter with a disability⁶ and the United States Department of Education has withdrawn numerous disability guidance documents.⁷ In March 2018, “[a] federal magistrate judge . . . held the Law School Admission Council [(LSAC)] . . . in contempt of court for partly violating a consent decree establishing procedures to handle requests for disability accommodations on the [LSAT].”⁸ If these trends continue, law school classrooms may soon be filled even more disproportionately with white, middle class, non-disabled, and overwhelmingly male students.

This is a multi-faceted problem with no simple solution. This Article addresses one small piece of this problem: the use of speeded exams as an aspect of the law school admissions process. Ending this practice would likely improve the diversity of law schools, while also helping admissions offices admit a more qualified class to practice law.

Stereotypically, faster is considered better.⁹ Starting in the early years of education, the most lauded students are those who do best on speeded¹⁰ standardized exams,¹¹ even though the scores on such exams do not correlate

Cheating?, CHRON. HIGHER EDUC. (Sept. 18, 2016), <https://www.chronicle.com/article/Extra-Time-on-an-Exam-/237787>.

⁶ See CNN, *Trump Mocks Reporter with Disability*, YOUTUBE (Nov. 25, 2015), <https://www.youtube.com/watch?v=PX9reO3QnUA>.

⁷ See Valerie Strauss, *Education Department Withdrawing Nearly 600 Policy Documents It Says Are Outdated*, WASH. POST (Oct. 27, 2017), https://www.washingtonpost.com/news/answer-sheet/wp/2017/10/27/education-department-withdrawing-nearly-600-policy-documents-it-says-are-outdated/?utm_term=.ba446a7e0d3b.

⁸ Debra Cassens Weiss, *Council That Administers the LSAT is Held in Contempt; ADA Consent Decree is Extended*, A.B.A. J. (Mar. 6, 2018), http://www.abajournal.com/news/article/council_that_administers_the_lsac_is_held_in_contempt_ada_consent_decree_is/.

⁹ For a critique of this common view, see Simon Garfield, *Faster Isn't Always Better—We Should All Try Living at a Slower Pace*, GUARDIAN (Sept. 1, 2017), <https://www.theguardian.com/books/2017/sep/01/simon-garfield-reformation-2017-faster-isnt-always-better>. Testing is not the only context in which an overdue emphasis on speed harms people with disabilities. For example, disability activist, lawyer, and mother, Carrie Ann Lucas, has written poignantly about the challenges she faces while using assistive communication technology because people are too impatient to allow her to type her words to communicate. See Carrie Ann Lucas, *Communication and Privilege*, DISABILITY PRIDE (May 18, 2018), <http://www.disabilitypride.com/2018/05/28/communication-and-privilege/>.

¹⁰ A “speeded exam” is one in which the time constraints require a test taker to move rapidly through the test questions so that one’s ability to work quickly is a part of a test taker’s final score. A speeded exam will always have time constraints that are intended to influence how a test taker moves through the exam questions. A pure speed exam would be one in which the questions are relatively easy so that the only element that is measured is the ability to move rapidly through questions. Most standardized exams, however, contain challenging questions so that they measure both speed and power. For further discussion of the difference between “speed” and “power,” see 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, 37 (2005).

¹¹ For example, National Merit recognition is only possible if a high school junior scores in approximately the top one percent of all test takers on a speeded, multiple-choice exam.

with a student's ability to solve difficult problems,¹² and do correlate in a perfect linear relationship with socio-economic status.¹³

While testing entities provide applicants an opportunity to demonstrate that they are disabled¹⁴ and should be provided with extended time to take standardized exams,¹⁵ these testing entities also stand firmly behind the speeded nature of the exams. Hence, testing entities impose significant time constraints on all exam takers, while making extended time available to a small group of applicants with disabilities through an onerous¹⁶ and expensive¹⁷ process of obtaining a diagnosis and completing extensive

See *National Merit Scholarship Program*, NAT'L MERIT SCHOLARSHIP CORP., <https://www.nationalmerit.org/s/1758/interior.aspx?sid=1758&gid=2&pgid=424> (last visited Nov. 21, 2018); see also *The SAT: Questions and Answers*, FAIRTEST, <http://www.fairtest.org/facts/satfact.htm> (last visited Nov. 21, 2018) (criticizing bias in the SAT and PSAT, including bias stemming from the speeded nature of the exams).

¹² Researchers have found that measures that emphasize speed correlate near zero, perhaps negatively, with tests that require solving difficult problems. See John L. Horn & A. Nayana Blankson, *Foundations for Better Understanding of Cognitive Abilities*, in CONTEMPORARY INTELLECTUAL ASSESSMENT: THEORIES, TESTS, AND ISSUES 73 (Dawn P. Flanagan & Patti L. Harrison eds., 3d ed. 2012). For a summary of the research indicating that speed is not a component of an individual's reasoning ability, 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, 979 (2004).

¹³ See COLL. BD., 2012 COLLEGE-BOUND SENIORS: TOTAL GROUP PROFILE REPORT 4 tbl.11 (2012) [hereinafter COLL. BD., 2012 COLLEGE-BOUND SENIORS], <http://media.collegeboard.com/digitalServices/pdf/research/TotalGroup-2012.pdf> (reporting higher average test scores on all three sections of the SAT for each \$20,000 family income increment).

¹⁴ In this Article, when the author refers to a student being "disabled," the author means that the student has been identified as disabled under the definition provided by the Americans with Disabilities Act (ADA). See 42 U.S.C. § 12102(1) (2018) ("The term 'disability' means, with respect to an individual (A) a physical or mental impairment that substantially limits one or more major life activities of such individual; (B) a record of such an impairment; or (C) being regarded as having such an impairment (as described in paragraph (3)).").

¹⁵ See *Accommodations that May Be Available on the LSAT*, LAW SCH. ADMISSION COUNCIL, INC., <https://www.lsat.org/lsat/lsac-policy-accommodations-test-takers-disabilities/accommodations-may-be-available-lsat> (last visited Nov. 21, 2018) [hereinafter LAW SCH. ADMISSION COUNCIL, INC., *Accommodations that May Be Available*] (listing types of accommodations that may be available to test takers, including extended time); *LSAC Policy on Accommodations for Test Takers with Disabilities*, LAW SCH. ADMISSION COUNCIL, INC., <https://www.lsat.org/jd/lsat/accommodated-testing> (last visited Nov. 21, 2018) [hereinafter LAW SCH. ADMISSION COUNCIL, INC., *LSAC Policy on Accommodations*] (LSAC testing accommodations policy).

¹⁶ For example, the College Board warns applicants that the approval process can take seven weeks. See *Submitting a Request*, C. BOARD, <https://www.collegeboard.org/students-with-disabilities/request-accommodations> (last visited Nov. 21, 2018). LSAC entered into a consent decree to resolve the onerous requirements it placed on applicants who sought accommodations. See *Consent Decree, Dep't of Fair Emp't & Hous. v. Law Sch. Admission Council Inc.*, No. CV 12-1830-EMC (N.D. Cal. May 29, 2014) [hereinafter LSAC Consent Decree], https://www.ada.gov/lsac_consentdecree.htm.

¹⁷ See, e.g., Abigail Sullivan Moore, *Accommodations Angst*, N.Y. TIMES (Nov. 4, 2010),

paperwork.

Some universities make submission of scores on the speeded ACT test¹⁸ or Scholastic Assessment Test (SAT)¹⁹ optional, allowing applicants to use other methods to demonstrate their knowledge and abilities.²⁰ Nonetheless, even when applicants take advantage of such optional methods, universities have to be concerned that applicants who provide test scores may have a competitive advantage with respect to admissions or financial aid, including merit-based financial aid.²¹ The *absence* of standardized test scores signals that an applicant's scores are well below the median for that university's applicant pool. These test-optional universities presumably stand firmly behind the importance of speeded standardized exams because they allow applicants to submit their scores on such exams as part of the application process.²² Further, because standardized test scores are part of the *U.S. News and World Report* ranking system, even test-optional universities need to be mindful of the test scores of their admitted students in order to attain a high-ranked status.²³

<https://www.nytimes.com/2010/11/07/education/edlife/07strategy-t.html> (reporting that private testing "can cost \$1,000 to \$5,000").

¹⁸ For information on the ACT test, see ACT, <https://www.act.org> (last visited Nov. 21, 2018).

¹⁹ For information on the SAT, see *SAT Suite of Assessments*, C. BOARD, <https://collegereadiness.collegeboard.org/sat> (last visited Nov. 21, 2018).

²⁰ See *More Than 1000 Accredited Colleges and Universities that Do Not Use ACT/SAT Scores to Admit Substantial Numbers of Students into Bachelor-Degree Programs*, FAIRTEST, <https://www.fairtest.org/university/optional> (last visited Nov. 21, 2018).

²¹ Historically, financial aid was synonymous with financial need. With the advent of merit-based financial aid, however, financial aid dollars have begun to flow dramatically to high-income students who accrue strong test scores. See Laura Grey, *What Test-Optional Means for College Admissions*, COLLEGEVINE: BLOG (July 11, 2018), <https://blog.collegevine.com/what-test-optional-means-for-college-admissions/> (reporting that students who do not submit test scores have lower admission rates and may be ineligible for certain kinds of merit aid); see also Kolodner, *supra* note 2.

²² Nonetheless, the success of students who are admitted at test-optional universities suggests that speeded, standardized testing is not needed for admission purposes. See generally CHARLES ROONEY & BOB SCHAEFFER, FAIRTEST, TEST SCORES DO NOT EQUAL MERIT: ENHANCING EQUITY & EXCELLENCE IN COLLEGE ADMISSIONS BY DEEMPHASIZING SAT AND ACT RESULTS (1998), <https://www.fairtest.org/sites/default/files/optrept.pdf>. Hampshire College, by contrast, will not permit applicants to submit test scores, signaling that it does not believe such scores have value in the admission process. See Valerie Strauss, *The List of Test-Optional Colleges and Universities Keeps Growing Despite College Board's Latest Jab*, WASH. POST (Apr. 12, 2017), https://www.washingtonpost.com/news/answer-sheet/wp/2017/04/12/the-list-of-test-optional-colleges-and-universities-keeps-growing-despite-college-boards-latest-jab/?utm_term=.4814368177a6.

²³ See Robert Morse, Eric Brooks & Matt Mason, *How U.S. News Calculated the 2019 Best Colleges Rankings*, U.S. NEWS & WORLD REP. (Sept. 9, 2018), <https://www.usnews.com/education/best-colleges/articles/how-us-news-calculated-the-rankings> (reporting that U.S. News includes the SAT or ACT scores of test-optional schools in its rankings).

At the graduate level, such as admission to law schools, applicants typically have no choice but to take a standardized exam for admissions purposes.²⁴ If meritorious applicants are not sufficiently disabled to attain extended time, but have a history of poor test taking, then they are simply disadvantaged during the application process. They are forced to plead with the admissions office to overlook their low test scores, despite their strong academic record or work experience, because the scores are presumptively considered a valid indicator of their merit. And these applicants are likely to do poorly in the competition for merit-based financial aid, in which standardized test scores are an important factor.

In addition to being onerous and expensive, these kinds of work-arounds—i.e., applying to universities that are test-optional, seeking extended time, or requesting that low test scores on exams taken under regular conditions be ignored—likely have a negative effect on applicants' self-esteem. Even after being admitted, students who took advantage of these work-arounds may wonder if they are as qualified as the rest of the student body. In short, students who would attain lower test scores under standard speeded conditions may buy into the merit myth even if they are admitted on the basis of these alternative paths.²⁵

²⁴ See, e.g., AM. B. ASS'N, SEC. OF LEGAL EDUC. & ADMISSIONS TO THE B., ABA STANDARDS AND RULES OF PROCEDURE FOR APPROVAL OF LAW SCHOOLS, Standard 503 (2017–2018), https://www.americanbar.org/content/dam/aba/publications/misc/legal_education/Standards/2017-2018ABAStandardsforApprovalofLawSchools/2017_2018_aba_standards_rules_approval_law_schools_final.authcheckdam.pdf [hereinafter ABA STANDARDS AND RULES OF PROCEDURE] (requiring a “valid and reliable admission test”). While this Standard is currently under consideration for revision, it is expected that nearly all law schools will continue to require the submission of the results on a standardized exam for admission purposes in the foreseeable future. See *infra* Part II.B.2. Because U.S. News & World Report gives significant weight to such scores for ranking purposes, law schools are unlikely to risk their rankings by failing to require submission of such scores. See Derek Muller, *The Sensational Hype Over Lawless Law School Admissions*, PRAWFSBLAWG (May 17, 2018), <http://prawfsblawg.blogs.com/prawfsblawg/2018/05/the-sensational-hype-over-lawless-law-school-admissions.html>.

²⁵ It has been well-documented that minority students often attain grades in law school that are lower than might be predicted by their LSAT score. One hypothesis for this discrepancy is the negative environmental influences that may cause minority students not to feel sufficiently confident when they take one timed, summative exam in a traditional law school class. See, e.g., John Fordyce et al., *Predicting First-Year Law School Performance: The Influences of Race, Gender, and Undergraduate Major*, 43 EASTERN ECON. J. 64 (2017); Alexia Brunet Marks & Scott A. Moss, *What Predicts Law Student Success? A Longitudinal Study Correlating Law Student Applicant Data and Law School Outcomes*, 13 J. EMPIRICAL LEGAL STUD. 205 (2016); Daniel Schwarcz & Dion Farganis, *The Impact of Individualized Feedback on Law Student Performance*, 67 J. LEGAL EDUC. 139 (2017).

This Article will seek to import test-validity principles under Title VII of the Civil Rights Act of 1964²⁶ in the employment context to the reasonable accommodation process under the Americans with Disabilities Act (ADA)²⁷ for standardized exams to argue that we should change the default rules regarding the use of speeded exams. Because speeded exams cause a disparate impact on the basis of disability,²⁸ entities should not be permitted to use such exams for admissions or selection purposes unless the entities can demonstrate that the speeded use of those exams is a valid predictor of what the exam is designed to measure *and* that no less-impactful alternative is available, such as a non-speeded exam,²⁹ which would also be a valid predictor of what the exam is designed to measure. That is the traditional rule under Title VII, which applies to employers who use standardized exams that have a disparate impact on the basis of race or gender when selecting candidates for hiring or promotion.³⁰

If the Title VII validity standard were imported into the ADA, it is unlikely that testing entities could meet these rigorous standards to validate the speeded nature of their exams. To date, no testing entity has published a valid study justifying the time limits that are imposed on their exams. Further, it is likely that a non-speeded exam would be at least as valid as the current speeded testing instruments. As a logical matter, it is unlikely that the ranked test scores on an exam where a large number of applicants did not even finish are more valid than the ranked test scores on an exam where

²⁶ See 42 U.S.C. § 2000e-2(h) (2018) (permitting an employer “to act upon the results of any professionally developed ability test provided that such test . . . is not designed, intended or used to discriminate because of race, color, religion, sex or national origin”).

²⁷ See *id.* § 12189 (requiring “examinations . . . related to applications, licensing, certification, or credentialing for secondary or post-secondary education” to be offered “in a place and manner accessible to persons with disabilities”).

²⁸ See, e.g., Nonie K. Lesaux et al., *The Effects of Timed and Untimed Testing Conditions on the Reading Comprehension Performance of Adults with Reading Disabilities*, 19 *READING & WRITING* 21 (2006).

²⁹ A “non-speeded exam” is a test in which all test takers are given ample time to read and closely consider the answer to all questions. See *supra* note 10 and accompanying text. It is not literally “unspeeded” in that it could have a reasonable end time. Thus, for example, if at least ninety-nine percent of the applicant pool can complete an exam in four hours, then it would be appropriate to make four hours the default time limit for all test takers. If someone had a disability-related reason for needing more than four hours, then that individual could make an individualized request. Because nearly all applicants with disabilities would be able to complete the exam in four hours, then the exam conditions would not produce a disparate impact on the basis of disability. But, under the argument made in this Article, the test developer would have to establish that a four-hour time limit permitted at least ninety-nine percent of the applicant pool to finish the test. It could not set the four-hour limit merely out of administrative convenience without knowing the impact of that time limit on the applicant pool.

³⁰ See *infra* Part III.

every applicant is given an opportunity to finish.³¹ The emphasis on rank-ordered scoring through a speeded testing instrument helps testing entities provide rankings at the top of the scale among applicants who are not meaningfully distinguishable on the basis of their test scores alone,³² at the expense of many applicants who do not even finish the exam.³³ As will be discussed in Part II, the only major testing entity to consider the validity of a non-speeded exam found that test validity was not lessened when the exam became non-speeded.

Considerable controversy exists over questions regarding which applicants are disabled and exactly how much extended time individuals with disabilities should obtain as an accommodation on standardized exams.³⁴ This Article argues that we are mired in asking the wrong questions. Instead of asking who is disabled and precisely how much extended time that applicant should receive, we should ask: Do speeded exams produce disparate impacts, and would the exams be equally valid if they were non-speeded? If so, then everyone should be given sufficient time

³¹ Researchers have suggested that additional time allows nearly all examinees to better demonstrate their knowledge because all test takers do better when given an opportunity to complete an exam and check their answers. See Cheryl L. Wild et al., *Effect of Increased Test-Taking Time on Test Scores by Ethnic Group, Years Out of School, and Sex*, 19 J. EDUC. MEASUREMENT 19 (1982).

³² Hence, in private conversations with the author, admission officers at top law schools insist that they look beyond test scores to distinguish among strong applicants because the very small percentile differences in their test scores are not meaningful or important in deciding whom to admit to a class.

³³ The raw score conversion chart for the LSAT supports this argument. Ten points (170 to 180) distinguishes applicants who rank between the 97.4 and 99.9 percentiles. See *LSAT Score Conversion*, ALPHA SCORE, <https://www.alphascore.com/resources/lsat-score-conversion/> (last visited Nov. 25, 2018) [hereinafter *LSAT Score Conversion*]. An applicant who scores a 170 can miss about fifteen percent of the exam. At the bottom end of the test instrument, a score of 120 to 130 distinguishes applicants who rank between the 0.1 percentile and 2.0 percentile. *Id.* An applicant who scores 130 can correctly answer 25 of 100 questions, as compared to an applicant who correctly answers 15 of 100 questions and receives a 120. Because there are five test item choices, random guessing should produce a raw score of twenty (one out of five correct answers), which converts to a 125 scaled score. *Id.* It is inconceivable that the difference in score between a 120 and 125 is meaningful, because a score of 125 results from having correctly answered twenty questions—the result that is most likely from random guessing. Maybe some of the students in the bottom range do not even know to guess; instead they leave questions blank. The important point is that it is not likely that the LSAT provides useful correlation data to distinguish a student who scored a 130 from a student who scored a 120 because the raw score is so close to the random guessing score. Very few students with scores between 120 and 130 even enroll in law school; accordingly, it is also hard to see how the test developers can create validity data for that group. But, as the author will discuss, guessing is not limited to those who score between 120 and 130. Even students who score in the above-average range report that they engage in significant guessing on the exam. See *infra* Parts II, V.

³⁴ Media stories on this subject are common. See, e.g., Belkin, *supra* note 5; Lewak, *supra* note 5; Trachtenberg, *supra* note 5.

to complete an exam under non-speeded conditions. Despite decades of angst over how exactly to decide whether to extend time for a subset of people with disabilities, there has been surprisingly little attention to whether speeded exams make sense for any test takers.³⁵ Further, no one has sought to make the argument that speeded standardized exams are often legally impermissible under the ADA.

To construct the argument that speeded exams are often legally impermissible, Part II will discuss the widespread use of speeded standardized exams, including the LSAT.³⁶ Because the American Bar Association (ABA) is in the process of considering revisions to its rules on standardized testing requirements,³⁷ this is an important moment to revisit the speeded aspect of the LSAT. Next, Part III will review the test validity rules under Title VII for employer-mandated exams, which require the validation of exams that produce race or gender-based disparate impacts. Part IV will contrast those rules with the ADA approach for standardized exams, which often applies to admissions testing and professional licensing, and which usually requires accommodation for some test takers rather than the validation of the aspect of the exam that produces disparate impacts. Finally, Part V will suggest how Title VII principles could be imported into the ADA to create a new legal standard requiring testing entities to demonstrate validity when there is a disability-based disparate impact, rather than require each individual student to request extended time as an alternative exam accommodation. This Part will also apply that argument specifically to the LSAT and deflect the common arguments that are made in favor of the speeded nature of the LSAT.

³⁵ *But see* William C. Kidder & Jay Rosner, *How the SAT Creates “Built-In Headwinds”: An Educational and Legal Analysis of Disparate Impact*, 43 SANTA CLARA L. REV. 131 (2002) [hereinafter Kidder & Rosner, *How the SAT Creates “Built-In Headwinds”*] (criticizing racial bias of testing instruments); 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 (2000) [hereinafter Kidder, *Portia Denied*] (criticizing gender bias of testing instruments); Henderson, *supra* note 12, at 1031–46 (arguing that the emphasis on time-pressured examination instruments has little theoretical connection to the actual practice of law and disadvantages candidates on the basis of race, gender, and age).

³⁶ For information on the LSAT, see LAW SCH. ADMISSION COUNCIL, INC., <https://www.lsac.org/index>. It has been estimated that seventy to eighty percent of admission decisions are determined exclusively on the basis of LSAT and undergraduate GPA. See Linda F. Wightman, *The Consequences of Race-Blindness: Revisiting Prediction Models with Current Law School Data*, 53 J. LEGAL EDUC. 229, 235 (2003).

³⁷ See Stephanie Francis Ward, *ABA Legal Ed Council Approves Proposed Rule Change to End Admission Test Requirement*, A.B.A. J. (May 11, 2018), http://www.abajournal.com/news/article/ABA_legal_education_council_rule_change_end_admission_test_requirementreq; *Council Adopts Proposal to Make Standardized Test Optional for Law Schools*, A.B.A. NEWS (May 14, 2018), https://www.americanbar.org/news/abanews/aba-news-archives/2018/05/council_adopts_propo.html.

Speeded testing is not just a disability problem. It poses competitive disadvantages for many applicants. A universal design solution,³⁸ under which all test takers are allowed to take an exam under non-speeded conditions, would benefit applicants with diagnosed and undiagnosed disabilities, as well as minority applicants, applicants with low socio-economic status, older applicants, and women.³⁹ Thus, although the universal design solution might technically emerge from the ADA, it would attain a range of diversity benefits. It would push testing companies to devise examination instruments that are not primarily designed to distinguish among the top 1% at the expense of a broad swath of diverse applicants. The choice of a speeded testing instrument is a choice about who is given an opportunity to fill the seats of our classrooms and at what price.

II. SPEED

A. Speeded Exams

1. Introduction

As a general matter, there are “speed” exams and “power” exams.⁴⁰ “[O]n a pure *speed* test, individual differences depend entirely on the speed of performance and test items are of relative ease, whereas on pure *power* tests, the differences are not based on speed and the items increase in difficulty.”⁴¹ Although most standardized exams contain test questions that

³⁸ “Universal design for learning (UDL) is a framework to improve and optimize teaching and learning for all people based on scientific insights into how humans learn.” *About Universal Design for Learning*, CAST, <http://www.cast.org/our-work/about-udl.html#.Wvsjzy-ZM9w> (last visited Feb. 13, 2019). In the Higher Education Opportunity Act of 2008, Congress provided the following definition of universal design for learning:

a scientifically valid framework for guiding educational practice that (A) provides flexibility in the ways information is presented, in the ways students respond or demonstrate knowledge and skills, and in the ways students are engaged; and (B) reduces barriers in instruction, provides appropriate accommodations, supports, and challenges, and maintains high achievement expectations for all students, including students with disabilities and students who are limited English proficient.

20 U.S.C. § 1003(24) (2018).

³⁹ See generally Henderson, *supra* note 12, at 997–99 (documenting the adverse impact of various standardized tests on the basis of age, gender, and race); Kidder, *Portia Denied*, *supra* note 35, at 29 (summarizing studies that indicate that women are more likely to omit questions than men on a speeded exam, and that African-Americans experience greater score gains than whites when time pressures are eased). But see Wild et al., *supra* note 31, at 27–28 (concluding that extending the existing time limits of standardized verbal and quantitative tests will not differentially increase scores of groups defined by race, sex, or years out of school, but suggesting further study on tests that have a “practice effect”).

⁴⁰ See Ofiesh et al., *supra* note 10, at 37.

⁴¹ *Id.* (emphasis in original) (citation omitted).

are not easy, and therefore are power tests, they also have significant time limits that create a speeded component.

When these standardized exams are taken with traditional time limits, the speeded component disadvantages many test takers with learning disabilities in reading because these individuals proceed more slowly than their non-disabled peers on speeded tasks.⁴² They do not get to demonstrate their “power” on the test questions that they do not complete, even though they may be accurate, although slow, readers.⁴³ Because of the seeming unfairness of penalizing a group of people who read more slowly due to a disability, test administrators began to provide mechanisms that would allow individuals with disabilities to seek extended time to take a standardized test so that they would have an equal opportunity to demonstrate their skills and abilities.⁴⁴ Such accommodations are also required by the ADA.⁴⁵ As Sally Shaywitz has observed: “[d]yslexia robs a person of time; accommodations return it.”⁴⁶

Nonetheless, the allocation of extended time to some students, but not all students, has also created a fair amount of controversy: Who should be entitled to extended time? How much extended time? Does extended time create fairness problems? As Professor Sireci and his co-authors have remarked, the literature on these questions is “vast and passionate.”⁴⁷ One can find news articles with headlines such as “*Rich Parents Are Using Doctor’s Notes to Help Kids Cheat the SATs*,”⁴⁸ which reflect the controversial nature of extended time.

The traditional response to these criticisms is that conventional time limits shift the power/speed balance for people with disabilities so that the exam becomes overly speeded. Conventional time limits impose “construct-irrelevant variance” by measuring someone’s disability rather than the constructs intended to be measured by the test items.⁴⁹

⁴² See generally SALLY SHAYWITZ, *OVERCOMING DYSLEXIA: A NEW AND COMPLETE SCIENCE-BASED PROGRAM FOR READING PROBLEMS AT ANY LEVEL* 29–31 (First Vintage Books ed. 2005) (discussing prevalence of dyslexia).

⁴³ See Ofiesh et al., *supra* note 10, at 37.

⁴⁴ See Stephen G. Sireci et al., *Test Accommodations for Students with Disabilities: An Analysis of the Interaction Hypothesis*, 75 REV. EDUC. RES. 457, 481 (2005).

⁴⁵ See 42 U.S.C. § 12189 (2018); 28 C.F.R. § 36.309 (2018).

⁴⁶ SHAYWITZ, *supra* note 42, at 314.

⁴⁷ Sireci et al., *supra* note 44, at 458.

⁴⁸ Lewak, *supra* note 5 (characterizing requests for extra time as “snowflake behavior”).

⁴⁹ See Sireci et al., *supra* note 44, at 458 (“Viewed in the perspective of test validity theory, some features of a standardized test administration introduce *construct-irrelevant variance* for some students.” (emphasis in original)).

Construct-irrelevance is easy to understand in the context of certain disabilities. If we handed a print exam to a blind student, we would be measuring the construct-irrelevant ability to read regular print rather than the skills or abilities the testing instrument is designed to measure. Similarly, if we handed a paper and pencil exam to a student who did not have use of his or her hands, we would be measuring the construct-irrelevant ability of fine motor skills rather than the skills or abilities the testing instrument is designed to measure. In these two examples, one can readily understand how certain test conditions distort the construct the exam is designed to measure, so that test accommodations become appropriate to ensure that the test measures the intended attributes rather than the person's impairment.

The construct-irrelevance argument, however, becomes more difficult to grasp when test takers with disabilities seek extended time to take a speeded standardized exam because the time limits pose challenges for many disabled test takers. By contrast, typical test takers do not struggle to see the print on an exam or use a pencil to record their answers. Typical test takers would not expect to attain any score benefit by being provided an auditory copy of the exam or the ability to use an instrument other than a pencil to record an answer. Typical test takers, however, would often perceive that they, too, would benefit from extended time.

The response to this argument is that conventional time limits disproportionately preclude people with disabilities from demonstrating the actual knowledge, skills, and abilities that the speeded exam is intended to test. The purpose of extended time is to allow people with disabilities to experience the same amount of speed constraint as non-disabled test takers. They are not supposed to be advantaged; the extended time merely levels the playing field so that everyone has an equivalent ability to demonstrate their actual knowledge, skills, and abilities. But, as will be shown, it is difficult to allocate the precise amount of extended time that will allow the disabled test taker to experience a similar speed constraint as typical test takers. The solution to this problem, however, is not to place more burdens on test takers with disabilities; the solution is to place more burdens on the testing entities who have access to huge swaths of data concerning their testing instruments by requiring them to validate the time limit that is imposed on all test takers. For now, however, we need to understand how the extended time accommodation was developed so that we can better understand its strengths and limitations.

2. Maximum Potential Thesis

M. Kay Runyan developed an early justification for what she called “extra time” (what is now is commonly called “extended time”)⁵⁰ for disabled test takers in 1991,⁵¹ and which has been subsequently labeled the “Maximum Potential Thesis.”⁵² This thesis contends that “nonlearning disabled students would not benefit from extra time because they are already working at their maximum potential under timed conditions. In contrast, many students with learning disabilities process information more slowly and do not operate at full potential unless given the extra time.”⁵³

The maximum potential thesis has significant appeal because it resolves complaints about test unfairness when some students, but not others, get extended time. If extended time provides little or no benefit to non-disabled students, then there would be no concern about some test takers qualifying for extended time on the basis of inappropriate disability documentation. If extended time only helps those with genuine disabilities, then test takers with questionable disabilities would accrue no advantage from an inappropriate request for extended time.⁵⁴

Runyan was able to support the maximum potential thesis in a setting where non-disabled students are typically able to complete a test within the regular time limits, but the students with learning disabilities are not. In her research study, “students were not allowed [to use extended time] to go back and change answers.”⁵⁵ Thus, the students who were able to finish under the regular timed conditions had no work to do when provided with extended time. She found that extended time did not benefit the non-disabled students because nearly all of them were able to complete the test within the required

⁵⁰ This Article uses the phrase “extended time” rather than “extra time” because the term “extra” implies something additional beyond what is appropriate. Within the standardized testing community, there has been a subtle move to use the phrase “extended time” rather than “extra time.” See, e.g., *Extended Time Accommodation*, C. BOARD, <https://www.collegeboard.org/students-with-disabilities/typical-accommodations/time> (last visited Oct. 27, 2018); LAW SCH. ADMISSION COUNCIL, INC., *Accommodations that May Be Available*, *supra* note 15.

⁵¹ See M. Kay Runyan & Joseph F. Smith, Jr., *Identifying and Accommodating Learning Disabled Law School Students*, 41 J. LEGAL EDUC. 317 (1991); 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) [hereinafter Runyan, *The Effect of Extra Time*].

⁵² G. E. Zuriff, *Extra Examination Time for Students with Learning Disabilities: An Examination of the Maximum Potential Thesis*, 13 APPLIED MEASUREMENT EDUC. 99, 101 (2000).

⁵³ *Id.*

⁵⁴ See Brent Bridgeman et al., *Impact of Fewer Questions Per Section on SAT I Scores*, 41 J. EDUC. MEASUREMENT 291, 291 (2004).

⁵⁵ Runyan, *The Effect of Extra Time*, *supra* note 51, at 107.

time limits.⁵⁶ By contrast, extended time allowed the students with learning disabilities to finish the test and, consequently, improve their score. With that kind of test design, it is no surprise that she found that only students with learning disabilities benefitted from extended time.

While some researchers have been able to replicate the maximum potential thesis,⁵⁷ others have criticized it as having a flawed research design because it did not allow students to use extended time to correct their answers.⁵⁸ Cheryl Wild and her co-authors conducted an important early study that explored the effect of extended time for all test takers on an exam with a speeded element under conditions in which test takers could use extended time to correct their answers.⁵⁹ Their focus was the effect of test taking time on the performance of women, older test takers, and racial minorities; disability was not their focus. They found that extended time did not differentially benefit any particular racial, age-based, or gender-based subgroup. They hypothesized that this result was due to the fact that “extra time allows examinees who have completed the test to review their answers.”⁶⁰ Their results have an intuitive resonance for exams that have a speeded component for all test takers because few test takers have an opportunity to complete all the questions to the best of their ability by carefully checking their responses. Additional time can improve everyone’s accuracy in that context, suggesting that the maximum potential thesis may be wrong to postulate that only test takers with disabilities benefit from extended time. Test takers with disabilities exclusively attain benefit from extended time only when the time limitations are so minimal that all non-disabled test takers have an ability to finish and check their answers within the prescribed time. Otherwise, the maximum potential thesis is on shaky ground.

3. Differential Boost Thesis

Because of the flaws with the maximum potential thesis, Stephen Sireci and his co-authors have concluded that a “differential boost” thesis makes more sense than the maximum potential thesis.⁶¹ They agree with the assertion, originally postulated by Wild, that “extra time appears to improve the performance of all student groups, not just those with disabilities,”

⁵⁶ *Id.*

⁵⁷ *See, e.g.,* Anthony J. Onwuegbuzie & Michael A. Seaman, *The Effect of Time Constraints and Statistics Test Anxiety on Test Performance in a Statistics Course*, 63 J. EXPERIMENTAL EDUC. 115, 120 (1995) (finding that time constraints disparately caused students with test anxiety to underperform in a statistics course).

⁵⁸ *See, e.g.,* Zuriff, *supra* note 52, at 106–08.

⁵⁹ *See generally* Wild et al., *supra* note 31.

⁶⁰ *Id.* at 26.

⁶¹ Sireci et al., *supra* note 44, at 481.

because “many educational tests are speeded to some extent.”⁶² Nonetheless, they concluded that the literature supports the observation that test takers with disabilities benefit *more* from extended time than non-disabled test takers, probably because they are using the extended time to complete their exam rather than to merely check their answers.

The problem, as observed by Sireci and his co-authors, is that there is no perfect way to determine precisely how much extended time a student with a disability needs to produce a test score that is equivalent to that of the non-disabled population.⁶³ The rationale for extended time is that people with disabilities have slower processing and reading speeds so they need additional time to demonstrate their skills and abilities. But, as Sireci and his co-authors note, “speed of processing is not a unitary concept and is at best a symbolic representation of a complex set of underlying cognitive schemes, actions, and skills that are the true source of the disability.”⁶⁴

Moreover, it is difficult to determine how much extended time is appropriate when test developers often assert that they do not intend to create a speeded exam.⁶⁵ If the exam were not testing speed, then the maximum potential thesis would apply—it would not matter how much extended time is allocated to test takers with disabilities because non-disabled test takers would be able to easily complete the exam in the allocated time. The non-transparency of the speeded component of standardized exams, however, makes it difficult for professionals to suggest how much extended time an individual test taker with a disability should receive on a particular examination instrument. The problem is not that professionals are sloppy or unprofessional; the problem is that the testing companies provide them with no information explaining the intended degree of speed constraint on the exam. Instead, as we will see in Part II with respect to LSAC, test companies do not justify why the exam needs to be speeded and then create impermissible barriers when test takers request extended time.⁶⁶

⁶² *Id.* at 483.

⁶³ *Id.* at 484 (“[T]here are no assessment techniques or instruments currently available that would allow us to make a precise ‘student by time needed’ judgment.”).

⁶⁴ *Id.*; see also Ofiesh et al., *supra* note 10, at 36 (noting that “a variety of speeded constructs exist, including reaction time, inspection time, decision speed, cognitive efficiency, and processing speed”).

⁶⁵ Sireci et al., *supra* note 44, at 483.

⁶⁶ See generally LSAC Consent Decree, *supra* note 16 (establishing rules for determining when test takers are entitled to use extended time accommodations). LSAC was later held in contempt of court for not complying with the consent decree. See *Dep’t of Fair Emp’t & Hous. v. Law Sch. Admission Council Inc.*, No. 12-CV-1830-JCS, 2018 WL 1156605, at *18–19 (N.D. Cal. Mar. 5, 2018).

By effectively denying requests for accommodation without following the procedures set by the Panel for internal review, external review, notification to the candidate of the decision and of the opportunity to

4. Flaws and Difficulties

In understanding the flaws and difficulties with the current system of allocating extended time, it is helpful to remember how this process of requesting extended time occurs. Students with disabilities are typically required to submit a report from an expert in the field of educational psychology who has diagnosed their disability and made a recommendation for extended time.⁶⁷ This recommendation is usually in the form of a blanket suggestion—50% extended time or 100% extended time. If the request meets appropriate professional standards, then the student will be given that amount of extended time on a variety of testing instruments, such as the SAT, ACT, GRE, LSAT, and college exams,⁶⁸ with no individualized assessment of the exact role of speed on each of these testing instruments.⁶⁹ The challenge is that there is no scientific way to determine precisely how much extended time a test taker will need on an exam, especially when that determination needs to be made in advance of the person taking the exam and the testing entities publish no reports explaining the role that speed plays in the examination instrument.⁷⁰ The testing entities' lack of transparency

appeal, and external review of any appeal, LSAC routinely violated the Consent Decree's requirement that it "implement the Best Practices" imposed by the Panel.

Id. at *20; *see also infra* Part II.B.

⁶⁷ *See, e.g., Accommodations and English Learner Supports for US Students*, ACT, <https://www.act.org/content/act/en/products-and-services/the-act/registration/accommodations.html> (last visited Oct. 27, 2018); *Accommodations for Test Takers with Disabilities or Health-Related Needs*, EDUC. TESTING SERV., https://www.ets.org/gre/revised_general/register/disabilities?WT.ac=rx28 (last visited Oct. 27, 2018); *Disability Documentation Guidelines*, C. BOARD, <https://www.collegeboard.org/students-with-disabilities/documentation-guidelines/disability-documentation> (last visited Oct. 27, 2018); *Documentation Requirements*, LAW SCH. ADMISSION COUNCIL, INC., <https://www.lsac.org/jd/sat/accommodated-testing/documentation-requirements> (last visited Oct. 27, 2018).

⁶⁸ For example, LSAC automatically provides a student with the same amount of extra time that they previously received on another standardized test, such as the SAT or ACT. *See* LAW SCH. ADMISSION COUNCIL, INC., *LSAC Policy on Accommodations*, *supra* note 15.

⁶⁹ A university professor, for example, might be told that an unnamed student will be given a certain amount of extra time on an exam that the professor is administering. No one asks to see the actual exam to learn whether the exam has a speeded element. For example, when one professor was offering a 28-hour take-home exam with a strict 3,000 word limit, the professor was told that a student would be receiving 50% extra time. The professor had deliberately written the exam under universal design principles so that no student would feel a time constraint to finish the exam. If someone had looked at the exam and seen the strict 3,000 word limit, then that person might have realized that extra time would not be beneficial. The student who received 50% extra time was not likely advantaged since the exam was not speeded; in fact, the student may have been disadvantaged by being expected to spend that length of time on one exam, and not spend that time preparing for other exams or writing papers for other courses.

⁷⁰ *See* Ofiesh et al., *supra* note 10, at 49 (recognizing that the literature on extended time as an accommodation is based on the "performance on a standardized reading-based multiple-

about the role of speed on their exams makes the job of the disability professional, who is expected to make an accommodation recommendation, virtually impossible.

Because of the difficulty with determining precisely how much time to provide individual students with disabilities, some researchers have published articles criticizing the precise amount of extended time recommended by disability professionals. For example, Lawrence Lewandowski and his co-authors published a study that sought to determine whether 25% extended time should be more appropriate than 50% or 100% extended time.⁷¹ For their study, they administered a 76-test item version of the reading comprehension subtest of the Nelson-Denny Reading Test to 107 students.⁷² These test items were intended to be completed in 15 minutes. They compared the number of test questions the students with learning disabilities could correctly answer when awarded 25%, 50%, or 100% extended time, with the number the non-learning disabled students could answer. They concluded that “25% extra time may suffice for the typical LD student, 50% extra would be more than what some students require, and 100% extra time would confer an unfair advantage for some students with LD.”⁷³

Although they recognized some limitations of their study, Lewandowski and his co-authors suggested that their results might be applicable to accommodations provided on speeded examinations, such as the SAT, ACT, LSAT, and GRE.⁷⁴ Further, they claimed (without citation to any authority) that the reading comprehension measure on the Nelson-Denny “is similar to sections found on high-stake exams such as the . . . LSAT.”⁷⁵

But their reasoning is deeply flawed because of the way they extrapolate from the relatively easy Nelson-Denny reading comprehension test to a speeded and difficult exam like the LSAT. Sample questions from the Nelson-Denny reading comprehension test are available online.⁷⁶ These questions appear to be at a much lower level of difficulty than the kinds of

choice test, and [that] therefore it is unclear how the[] results would generalize to the amount of time needed for essay exams or other test formats”).

⁷¹ Lawrence Lewandowski et al., *Effects of Extended Time Allotments on Reading Comprehension Performance of College Students with and Without Learning Disabilities*, 31 J. PSYCHOEDUCATIONAL ASSESSMENT 326, 333 (2013).

⁷² *Id.* at 328–29.

⁷³ *Id.* at 333–34.

⁷⁴ *Id.* at 334–35.

⁷⁵ *Id.* at 335.

⁷⁶ See N.C. WILDLIFE RES. COMM’N, NELSON-DENNY READING TEST: PRACTICE QUESTIONS 1–2 (2017), https://www.ncwildlife.org/Portals/0/Careers/Documents/Nelson_Denny_Study_Guide.pdf.

questions found, for example, on the LSAT, which are also available online.⁷⁷ The Nelson-Denny reading comprehension sample exercise has a four-paragraph essay followed by five multiple-choice questions. To answer these questions, one has to identify a particular sentence, and find a paraphrase or definition for a word in that sentence. For example, the reader would have to know that a “peninsula” is a “body of land with water on three sides.”⁷⁸ On the sample LSAT provided by LSAC on its website, only one of the four sections could be considered even minimally equivalent to the Nelson-Denny reading comprehension test, in that it involves reading passages and answering questions. The first reading passage is a four-paragraph essay followed by eight multiple-choice questions. But the questions do not merely ask the reader to paraphrase or define a term or phrase in the passage. Instead, they ask the reader to make predictions or inferences.⁷⁹ The second reading passage involves even higher-level thinking because it asks the reader to compare and contrast two different passages through eight multiple-choice questions.⁸⁰

Thus, if, hypothetically, most students with learning disabilities need twenty-five percent extended time to attain an appropriate score on a relatively easy fifteen-minute Nelson-Denny reading test, it is very difficult to extrapolate as to how much extended time they would need on a dense and difficult four-hour exam, such as the LSAT. One could easily imagine that their fatigue would increase exponentially rather than linearly as the four-hour exam continues. On the power/speed continuum, the two tests are very different. The extended time to attain an appropriate score on the Nelson-Denny (which is largely a power test) is likely to be far less than the extended time needed to attain an appropriate score on the LSAT (which is also more of a speed test). It is astonishing that researchers, such as Lewandowski, generalize from an experimental administration of a fifteen-minute section of the relatively easy Nelson-Denny to a four-hour high-stakes exam.

The implicit take-away from the Lewandowski study is that students with disabilities are greedy—they are asking for too much accommodation on the basis of too little information. Thus, it is not surprising that Lewandowski often concludes, when hired by testing entities, that they

⁷⁷ See, e.g., LAW SCH. ADMISSION COUNCIL, INC., THE OFFICIAL LSAT PREPTEST: JUNE 2007 1–30 (2007) <https://www.lsac.org/sites/default/files/legacy/docs/default-source/jd-docs/sampleptjune.pdf>.

⁷⁸ See N.C. WILDLIFE RES. COMM’N, *supra* note 76, at 2 (Reading Comprehension, Question 2). The questions require no inferential or higher-level thinking.

⁷⁹ See *Sample Law School Admission Test: Section 4*, LAW SCH. ADMISSION COUNCIL, INC., <https://www.lsac.org/free-sample-test/section-4.asp> (last visited Nov. 25, 2018) (question 7 asks about an “inference” and question 8 asks about a “prediction”).

⁸⁰ *Id.*

should deny students' requests for accommodations.⁸¹ And those kinds of research studies are used to support Wall Street Journal headlines proclaiming "*Colleges Bend the Rules for More Students, Give Them Extra Help*,"⁸² which, in turn, cite school district lawyers who are opposed to testing accommodations for students.⁸³ Nonetheless, there is no scientific support for the assertion that test takers who receive extended time are over-accommodated; the reasoning in even the best studies, such as the Lewandowski study, is flawed.⁸⁴

5. Response to Flaws and Difficulties

Despite the flaws in how Lewandowski and his co-authors reported their work, their raw data can be used to support the argument that *the best way to assess a student's reading comprehension is through a non-speeded instrument*. This Article will revisit their study to see why it supports that conclusion.

The Lewandowski study involved the administration of the Nelson-Denny reading comprehension test to 26 students with learning disabilities ("LD") and 50 non-disabled students ("non-LD").⁸⁵ They recorded how many correct questions the students answered after 15, 22.5, and 30 minutes.⁸⁶ They compared the scores of the LD students who were granted extended time with the scores of the non-LD students who were limited to 15 minutes, and concluded that the students with LD over-performed when granted more than 50% extended time.⁸⁷ But they also reported how many

⁸¹ See LSAC's Brief in Opposition to DFEH's Motion for Civil Contempt Order Or, in the Alternative, Modification of the Consent Decree at 56–66, Dept. of Fair Emp't and Hous. v. Law Sch. Admission Council Inc., No. 12-CV-1830-JCS (N.D. Cal. Jan. 12, 2018) (containing an example of a Lewandowski denial decision for Examinee No. 37100345).

⁸² Douglas Belkin, *Colleges Bend the Rules for More Students, Give Them Extra Help*, WALL ST. J. (May 24, 2018), <https://www.wsj.com/articles/colleges-bend-the-rules-for-more-students-give-them-extra-help-1527154200>.

⁸³ See Miriam Kurtzig Freedman, *Wall Street Journal Quotes Me About the Overuse of Accommodations! When Will It Ever End?*, SCH. L. PRO (May 26, 2018), <http://schoollawpro.com/wall-street-journal-quotes-me-about-the-overuse-of-accommodations-when-will-it-ever-end/> (commenting on the use of her quote in a *Wall Street Journal* article).

⁸⁴ Other studies are even more flawed, showing no understanding at all of how the extended time accommodation process even works. See, e.g., Joshua F. Drake, *Disabling Academic Standards: Learning Disabilities and Time-and-a-Half Testing*, 31 ACAD. QUESTIONS 304 (2018) (claiming, without citation, that "learning disabled students receive . . . up to three times the allotted time . . . to take their tests at college" and criticizing the policy of not flagging test scores taken under conditions of accommodation) (author is a professor of music and humanities at Grove City College with no apparent qualifications in law or disability).

⁸⁵ Lewandowski et al., *supra* note 71, at 328, 335.

⁸⁶ *Id.* at 329–30.

⁸⁷ *Id.* at 333.

answers, of answers attempted, each group answered correctly within those time limits. In other words, they reported the students' accuracy rate for attempted questions. Further, they reported how many questions the LD students, as compared to the non-LD students, were able to reach at the various time limits.⁸⁸

The focus of the Lewandowski study was to ask how many correct answers the non-LD population was able to answer correctly after, for example, 15 minutes, compared with how many correct answers the LD population was able to answer correctly after, for example, 30 minutes (100% extended time).⁸⁹ If the LD students would attain a higher score than the non-LD students after being awarded extended time, then they would conclude that the LD students were over-accommodated. But their data can also be used for a different purpose: to provide insight into the benefits of the universal design principle of *no* time limits. Their data support the validity of using no time limits to measure the reading comprehension ability of LD students in comparison with non-LD students.

In order to assess the reading comprehension ability of the two groups, without regard to speed, we could ask *what is their accuracy rate for the questions they reached*, rather than worry about how many questions they actually answered. Table 1 reports the reading accuracy of the learning disabled and non-learning disabled groups for the questions they answered in the time allocated.

Table 1: Reading Accuracy Rate for Learning Disabled and Non-Learning Disabled Students at Various Time Limits

	LD Group	Non-LD Group
15 minutes	83.5 %	88.1 %
22.5 minutes	84.2 %	87.4 %
30 minutes	83.8 %	86.8 %

In other words, if speed is taken out of the equation by only assessing students on what they actually answered, the profile of the LD students looks only slightly worse than that of the non-LD students. And that result is what we might expect because it also turns out that the LD and non-LD students were not identical in their academic performance. The LD students reported that their college GPA was slightly lower than that of the non-LD students (2.91 as compared to 3.28).⁹⁰ Thus, the reading comprehension accuracy

⁸⁸ *Id.* at 331–32 tbls.1, 3.

⁸⁹ *Id.*

⁹⁰ *Id.* at 329.

score (when speed was not a factor) seemed to accurately predict how the LD students performed in college—about three or four percentage points below the non-LD students.

A non-speeded testing instrument could therefore accurately provide information about the reading comprehension scores of both the LD and non-LD students. We do not have to score an exam on the basis of an absolute number of questions that each student is expected to reach. Instead, we can look at the accuracy of the student for the questions that are, in fact, attempted. By making an exam non-speeded, we could also require all students to attempt the same number of questions.

While Table 1 reflects that the Nelson-Denny test can provide an accurate assessment of the comparative reading abilities of the LD and non-LD populations when time limits are eliminated, the Lewandowski data can also provide strong support for why it is so essential to find a method to fairly measure the reading abilities of the LD population. Lewandowski and his colleague's focus is to critique the allocation of fifty percent extra time as over-accommodating LD students. But their data can also provide very strong support for why *no* extended time is unfair to the LD students. Because testing entities frequently challenge the right of LD students to get *any* extended time, this kind of data is also important and often not available.

Let us assume that these students were given a timed testing instrument that contained the number of questions typically reached by non-LD students. The students are told to answer as many questions as possible within the time limits and to use guessing behavior for questions they cannot reach. We record their percentage of correct answers by putting the number of correct answers in the numerator, and the overall number of questions in the denominator. Further, we will assume there are five multiple-choice options for each question and students could correctly guess the right answer for twenty percent of the questions they do not reach.

Because Lewandowski and his co-authors reported how many questions the LD and non-LD students answered at the end of each time limit, we can use the number of questions reached by the non-LD students in the denominator as the actual number of test questions. For the numerator for the LD students, we can use the number of actual correct answers plus twenty percent of the questions that the student did not reach, under the assumption that they would have correctly guessed the answer to twenty percent of those remaining questions. Using those assumptions, the following table compares the reading comprehension score of the LD students with the reading comprehension score of the non-LD students at each of the three timed intervals.

Table 2: Reading Comprehension Scores of LD and Non-LD Students if No Extended Time to LD Students

	LD Group	Non-LD Group
15 minutes	69.1 % ⁹¹	88.1 %
22.5 minutes	71.1 % ⁹²	87.4 %
30 minutes	71.9 % ⁹³	86.8 %

As shown in Table 2, when the test was administered on a timed basis, the LD students performed much worse than we would expect in light of their college GPAs. Their reading comprehension scores now appear to be about fifteen to nineteen percent worse than that of the non-LD population, while their GPAs were only a few points lower than that of the non-LD population. These results suggest that their reading comprehension scores are around seventy percent, even though their scores were around eighty-three percent on a non-speeded testing instrument, as shown in Table 1. Although people might disagree about how much extended time should be provided to the LD population, these results also demonstrate the importance of *some* remedy under the ADA in order for speeded tests not to dramatically disadvantage many individuals with disabilities. Traditionally, that remedy has been extended time.

Instead of using the Lewandowski data to criticize the “rampant and arbitrary”⁹⁴ assignment of extended time accommodations, we can therefore use their data to challenge the random and arbitrary establishment of speeded conditions.⁹⁵ The *non-speeded* reading test for *all* test takers seems to do the

⁹¹ The LD group, on average, correctly completed 19.35 items; they completed 6.79 fewer items than the non-LD group. Lewandowski et al., *supra* note 71, at 331 tbl.1. The author assumed that the LD group would correctly answer 20% of those items, which would add 1.35 to their score. Thus, they would correctly complete 20.70 items out of 29.94 items (the number of items attempted by the non-LD group). That would result in an overall accuracy rate of 70.3% for the LD group after 15 minutes.

⁹² The LD group, on average, correctly completed 31.04 items; they completed 9.38 fewer items than the non-LD group. *Id.* The author assumed that the LD group would correctly answer 20% of those items, which would add 1.88 to their score. Thus, they would correctly complete 32.92 items out of 46.26 items (the number of items attempted by the non-LD group). That would result in an overall accuracy rate of 71.1%.

⁹³ The LD group, on average, correctly completed 42.08 items; they completed 11.51 fewer items than the non-LD group. *Id.* The author assumed that the LD group would correctly answer 20% of the remaining items, which would add 2.30 points to their score. Thus, they would correctly complete 44.38 of 61.74 items (the number of items attempted by the non-LD group). That would result in an overall accuracy rate of 71.9%.

⁹⁴ *Id.* at 335.

⁹⁵ Nonetheless, Lewandowski and his colleagues end their article with recognition of the importance of a universal design approach under which all students are provided adequate time to complete a test: “It is time for testing entities to initiate research on the importance (or

best job of giving us useful information about everyone's reading comprehension skills. Unfortunately, entities that offer standardized exams for admission to university, graduate school, or entry into a profession often aggressively litigate requests for extended time rather than consider why they need the speeded conditions at all.⁹⁶ The Lewandowski data provide strong evidence of the disparate impact of a speeded testing instrument and the comparative fairness of a non-speeded instrument. Because establishing the precise amount of extra time that each test taker should receive is difficult, the non-speeded version is likely to be the most valid.

6. A Promising Model

An exception to the traditional stance of aggressively defending time limits is the approach taken by Pearson, a company that designs and markets achievement tests for the K–12 population.⁹⁷ Pearson commissioned a study in 2002 to determine if their K–12 testing on the Stanford 10 Achievement Test (“Stanford 10”) would be equally valid under speeded and non-speeded conditions.⁹⁸ This study was conducted after Pearson had already relaxed the stated time limits on the exam so that the vast majority of students were able to complete it under regular time limits.⁹⁹ In other words, they had already tried to make their exam non-speeded.

Their study reported two important findings. First, they found “that the allowance of extended times accommodates disabled students so that they

lack thereof) of speed in performance; if speed is truly of no importance, time limits should be liberalized for all examinees, and if speed has any relevance, time limits should be empirically determined.” *Id.*

⁹⁶ See, e.g., *Turner v. Nat'l Council of State Bds. of Nursing*, 561 Fed. App'x. 661 (10th Cir. 2014) (professional nursing exam); *Enyart v. Nat'l Conference of Bar Exam'rs, Inc.*, 630 F.3d 1153 (9th Cir. 2011) (state bar exam); *Black v. Nat'l Bd. of Med. Exam'rs*, 281 F. Supp. 3d 1247 (M.D. Fla. 2017) (professional medical exam); *Dep't of Fair Emp't & Hous. v. Law Sch. Admission Council Inc.*, 896 F. Supp. 2d 849 (N.D. Cal. 2012) (LSAT); *Shaywitz v. Am. Bd. of Psychiatry & Neurology*, 848 F. Supp. 2d 460 (S.D.N.Y. 2012) (professional medical exam); *Mahmood v. Nat'l Bd. of Med. Exam'rs*, No. 12-1544, 2012 WL 5364689 (E.D. Pa. Oct. 31, 2012) (professional medical exam); *Elder v. Nat'l Conference of Bar Exam'rs*, No. C 11-00199 SI, 2011 WL 672662 (N.D. Cal. Feb. 16, 2011) (state bar exam); *Hoppe v. Coll. of Notre Dame of Md.*, 835 F. Supp. 2d 26 (D. Md. 2011) (comprehensive exams); *Falchenberg v. N.Y. State Dep't of Educ.*, 642 F. Supp. 2d 156 (S.D.N.Y. 2008) (state teacher certification exam).

⁹⁷ See *Learning Without Limits*, PEARSON, <https://www.pearson.com/us/> (last visited Feb. 13, 2019).

⁹⁸ See THOMAS E. BROOKS ET AL., PEARSON EDUC., INC., ASSESSMENT REPORT: TIMED VERSUS UNTIMED TESTING CONDITIONS AND STUDENT PERFORMANCE 5 (rev. ed. 2004), https://images.pearsonassessments.com/images/tmrs/tmrs_rg/TimedUntimed.pdf?WT.mc_id=TMRS_Timed_Versus_Untimed_Testing.

⁹⁹ *Id.* at 10 (“[P]rocedures established by Pearson over the past 80 years for setting time limits have allowed adequate time for non-disabled students to complete the test without undue strain or errors owing to time pressure.”).

may demonstrate what they have learned while not unfairly inflating the scores of non-disabled students.”¹⁰⁰ Second, they found that testing conditions did not become unduly long when the time limits were relaxed. In at least ninety-five percent of the classrooms, teachers did not need to allow more than twenty additional minutes for everyone to have an opportunity to complete the test.¹⁰¹ Due to the results of this study, Pearson stopped imposing time limits on their Stanford 10 exam.

The Pearson study is an important addition to the literature on speed constraints because it was conducted under real testing conditions rather than in an artificial laboratory environment. For an achievement exam, which was not intended to have a speeded component, it appears that there is no reason to provide time constraints. Rather than try to justify a particular time limit, or the standards for determining who is disabled, they asked whether time limits were needed at all. They found that eliminating time limits did not harm the validity of the testing instrument and its administrative feasibility.

None of the professional testing organizations, which administer tests for college or graduate admission, have published any studies on the validity of the time limits they impose on these exams, and whether the exams would be equally valid if time limits were relaxed for all test takers. It is not appropriate to extrapolate from one study on a K–12 population for an academic achievement test to the entire world of high-stakes standardized testing. But the Pearson study shows that testing entities can devote resources to asking those questions and then administer tests consistently with the empirical results.

B. *LSAT as a Speeded Exam*

1. Introduction

The LSAT is only one standardized exam that would be subject to the legal arguments raised in this Article. Because the use of this instrument as the sole admission exam for entry to law school is being reconsidered, this is an excellent moment to consider the role of its speeded aspect.

First, let us establish the premise that the LSAT is a speeded exam. The available evidence suggests that the LSAT is a speeded exam, even though LSAC has never directly acknowledged that fact.¹⁰² When LSAC published

¹⁰⁰ *Id.*

¹⁰¹ *Id.* The length of this exam is not clear. It is possible that the test items were supposed to take only ten minutes to complete so that twenty additional minutes constituted what may be thought of as “triple time.”

¹⁰² Although LSAC has never publicly acknowledged that the LSAT has a speeded component, it now makes test prep materials available for free on its website, which LSAC claims will help a student increase his or her speed in taking the exam. *See Official LSAT*

a report in 1994 suggesting that the exam was not speeded, it discounted the importance of answering every question with care. Sections were not considered speeded even when the evidence suggested that twenty percent of test takers did not reach the last item of a section.¹⁰³ LSAC researchers viewed that as a situation where eighty percent *finished* the exam instead of as a situation where twenty percent *did not finish* the exam despite a lack of penalty for guessing. By contrast, another study reported that more than half of LSAC test takers (fifty-three percent) found it necessary to engage in “rapid-guessing behavior” in order to complete the exam.¹⁰⁴

Because LSAC’s reports relating to whether students completed test items are somewhat dated and incomplete, the author worked with her colleague Professor Guy Rub to determine whether Moritz College of Law students remembered the LSAT as being a speeded exam. Professor Rub asked students in his first-year contracts class to use a clicker to report their recollection about how many questions they had to guess without reading the questions on the LSAT. He was careful to distinguish between guessing when a student did not know the answer and guessing when a student did not have time to read and consider a question.

These were the results:

- 41%: answered all the questions after reading them (even if they had to guess a few),
- 31%: ran out of time and had to guess up to 4 questions without reading the questions and answers,
- 7%: ran out of time and had to guess 5 to 10 questions without reading the questions and answers,
- 4%: ran out of time and had to guess more than 10 questions without reading the questions and answers,
- 10%: did not take the LSAT, and
- 6%: did not remember.¹⁰⁵

Prep, KHAN ACAD., https://www.khanacademy.org/prep/lsat?utm_source=lsac-site&utm_medium=lsac&utm_campaign=lsac-launch2018&utm_term=homebanner (last visited Nov. 25, 2018) (quoting a user of the Khan Academy’s LSAT study materials, who stated that the materials helped the student “get[] faster at understanding what the question [was] asking”). William Henderson has persuasively argued that the LSAT has a significant speeded component. See Henderson, *supra* note 12, at 980 (“[R]ecent research has suggested that the LSAT may be much more ‘speeded’ than originally believed.”).

¹⁰³ See LYNDA M. REESE & RUTH ANNE COTTER, LAW SCH. ADMISSION COUNCIL, INC., A COMPENDIUM OF LSAT AND LSAC-SPONSORED ITEM TYPES 1948–1994, at 15 (1994).

¹⁰⁴ Henderson, *supra* note 12, at 992 (reporting the results of a study conducted by Schnipke and Scrams).

¹⁰⁵ Thanks to Professor Guy Rub for submitting this survey to our students.

If one removes from the sample the students who did not take the LSAT or did not remember their experience with guessing, we have the following results:

- 48.8% completed the test without guessing,
- 36.9% guessed on up to 4 questions,
- 8.3% guessed on between 5 and 10 questions, and
- 4.7% guessed on more than 10 questions.¹⁰⁶

Given the impact of even a 3-point difference in one's raw test score on admission and financial aid, one can see how the speed component of the test had a potentially large impact on more than half of the test takers—and certainly 13% of the test takers.¹⁰⁷ Despite the difficulty with using someone's recollections, these results were remarkably similar to those from another researcher who found that only 47% of students did not find it necessary to engage in any rapid guessing.¹⁰⁸ This is more guessing behavior than LSAC has acknowledged exists, possibly because LSAC has difficulty knowing how much guessing has taken place when a student does record an answer for each question on its pencil and paper exam.¹⁰⁹ Despite the role of guessing on the LSAT, LSAC has reported no studies validating the time limits.

This guessing behavior has significant consequences to the students who are not able to complete the exam without guessing. At Moritz, for example, the students who completed Professor Rub's survey scored, on average, around 160 on the LSAT.¹¹⁰ On a 100 question LSAT, these

¹⁰⁶ To arrive at these percentages, the author assumed that there were 100 students in the class. If there were 100 students, the assumption would be that 16 of the 100 students provided no useful data because they did not take the test or did not remember it. Then, of the 84 students who provided answers, 41 of 84 (48.8%) completed the test without guessing, 31 of 84 (36.9%) guessed on up to 4 questions, 7 of 84 (8.3%) guessed on between 5 and 10 questions, and 4 of 84 (4.7%) guessed on more than 10 questions.

¹⁰⁷ These data reflect how important it is to acknowledge the significance of even a small amount of speed constraint on the LSAT. See *LSAT Score Conversion*, *supra* note 33. The LSAT has about 100 questions. *Id.* A student who guesses on only 10% of the questions would guess on 10 questions, 8 of which he or she is likely to miss. Missing 8 questions has a very significant impact on one's percentile rank on the exam. The 50th percentile is around a score of 151 or a raw score of 55. *Id.* A raw score of 47 is a 146 (29.5%) and a raw score of 63 is a 156 (67.4%). *Id.* An 8 point raw score differential therefore has a very dramatic effect on one's score, and should not be discounted by saying that the exam is not speeded if people can complete up to 90% of the exam.

¹⁰⁸ See Henderson, *supra* note 12, at 992 (reporting the results of a study conducted by Schnipke and Scrams).

¹⁰⁹ LSAC Executive Director, Kellye Testy, has informed the author that LSAC is moving toward a computer-generated version of the LSAT. That form of the exam will make it easier to determine how much guessing behavior is occurring on the exam, because LSAC will be able to collect data on how much time a test taker spent on each question and how many questions were answered during the last minute of the exam.

¹¹⁰ See generally OHIO STATE UNIV., STANDARD 509 INFORMATION REPORT 1 (2017),

students, on average, missed about 30 questions.¹¹¹ If a student had answered 3 more questions correctly, his or her score would have risen to about 162 (75th percentile for enrolled students);¹¹² if a student had answered 3 more questions incorrectly, his or her score would have fallen to about 158 (near the 25th percentile for enrolled students).¹¹³ Thus, small differences in the number of correct answers can have a significant impact on admissions and merit financial aid, because they can either push a student from the 50th to the 75th percentile of the applicant pool, or from the 50th to the 25th percentile.¹¹⁴ It is well known that most law schools provide merit financial aid, in part, on the basis of LSAT scores.¹¹⁵ Thus, the speeded nature of the exam has significant implications for most of the test taking population. Nonetheless, as will be argued in Part V, the speeded aspect of the exam has never been validated.

2. Current LSAT Discussion

In the extensive discussion that is taking place regarding whether law schools should be required to have applicants take the LSAT or another standardized test, no discussion has been given to the question of whether the LSAT or another standardized test¹¹⁶ should be modified so that they are no longer speeded.

<https://moritzlaw.osu.edu/about/wp-content/uploads/sites/4/2013/06/Std509InfoReport-119-119-12-14-2017-07-52-55.pdf> [hereinafter OHIO STATE UNIV., STANDARD 509].

¹¹¹ See *LSAT Score Conversion*, *supra* note 33.

¹¹² See OHIO STATE UNIV., STANDARD 509, *supra* note 110.

¹¹³ *Id.*

¹¹⁴ Moritz College of Law's website reports that, for the entering class of 2017, the 25th percentile LSAT score was 157 and the 75th percentile score was 162. *J.D. Admissions*, OHIO ST. U., <http://moritzlaw.osu.edu/admissions/jd/> (last visited Nov. 25, 2018). The required ABA disclosures indicate that the 50th percentile LSAT score was 160. OHIO STATE UNIV., STANDARD 509, *supra* note 110, at 1.

¹¹⁵ Anecdotally, the author has had law school applicants tell her that their law school told them that their financial aid award would increase, after they were already admitted, if they re-took the LSAT and raised their score. Because the applicants had already been admitted, the law school had no doubt about their qualifications to be successful. Presumably, the law schools wanted the applicants to have higher LSAT scores so as to raise their U.S. News & World Report profiles.

¹¹⁶ The other test that is now being used by some law schools in the admission process is the GRE. Like the LSAT, the GRE has strict time limits. See *What to Expect*, EDUC. TESTING SERV., https://www.ets.org/gre/revised_general/test_day/expect/ (last visited Nov. 25, 2018) (stating that the GRE is about three hours and forty-five minutes). Any argument made in this Article about the LSAT would be equally applicable to the GRE. Because of the longstanding use of the LSAT as an admissions tool, however, this Article focuses on the LSAT. For a list of law schools that use the GRE, see *infra* note 120.

A brief discussion of the LSAT as a mandatory admissions requirement will set the stage for a discussion of the current controversy regarding the mandatory use of the LSAT during the admissions process.¹¹⁷ The ABA's longstanding Standard 503 states:

A law school shall require each applicant for admission as a first-year J.D. degree student to take a valid and reliable admission test to assist the school and the applicant in assessing the applicant's capability of satisfactorily completing the school's program of legal education. In making admissions decisions, a law school shall use the test results in a manner that is consistent with the current guidelines regarding proper use of the test results provided by the agency that developed the test.¹¹⁸

The ABA has concluded that the LSAT is valid and that law schools can seek to validate other tests.¹¹⁹ In recent years, some law schools have sought to validate the GRE as an admissions test.¹²⁰ Although the ABA does not prescribe the weight given to the LSAT in the admissions process,¹²¹ most law schools give it significant weight. The *U.S. News & World Report* weighs the median LSAT score of enrolled applicants at 12.5% of a law school's total score in determining law school rankings. That weighting puts

¹¹⁷ For a general discussion of the LSAT, see Ruth Colker, *Extra Time as an Accommodation*, 69 U. PITT. L. REV. 413, 425–33 (2008).

¹¹⁸ See ABA STANDARDS AND RULES OF PROCEDURE, *supra* note 24, Standard 503.

¹¹⁹ *Id.* at Interpretation 503-1.

A law school that uses an admission test other than the Law School Admission Test sponsored by the Law School Admission Council shall demonstrate that such other test is a valid and reliable test to assist the school in assessing an applicant's capability to satisfactorily complete the school's program of legal education.

Id.

¹²⁰ See, e.g., *Law Schools that Accept GRE Scores for Their J.D. Programs*, EDUC. TESTING SERV., https://www.ets.org/gre/revised_general/about/law/ (last visited October 30, 2018). Law schools that use the GRE include:

Brigham Young University J. Reuben Clark Law School, Brooklyn Law School, Chicago-Kent College of Law, Columbia Law School, Florida State University College of Law, George Mason University Antonin Scalia Law School, Georgetown University Law Center, Hamd Bin Khalifa University Law School, Harvard Law School, John Marshall Law School, Northwestern University Pritzker School of Law, Pace University Elisabeth Haub School of Law, St. John's University School of Law, Texas A&M University School of Law, University of Arizona James E. Rogers College of Law, University of Hawai'i at Manoa William S. Richardson School of Law, Wake Forest University School of Law, Washington University School of Law, and Yeshiva University Benjamin N. Cardozo School of Law.

¹²¹ See ABA STANDARDS AND RULES OF PROCEDURE, *supra* note 24, Interpretation 503-2 (“This Standard does not prescribe the particular weight that a law school should give to an applicant's admission test score in deciding whether to admit or deny admission to the applicant.”).

pressure on law schools to also give the LSAT a significant emphasis in admissions and financial aid decisions to improve their ranking.¹²² Professors Michael Sauder and Wendy Espeland report “that merit-based scholarships now take primacy over scholarships based on need” in order to improve LSAT scores and, consequently, law school rankings.¹²³ That emphasis, in turn, has a negative effect on socio-economic diversity because it is well-known that standardized test scores correlate with socio-economic status.¹²⁴ In other words, scholarship money often goes to the least needy applicants.¹²⁵

On February 10, 2016, the University of Arizona James E. Rogers College of Law announced that it would begin to accept either the GRE or the LSAT for admissions purposes.¹²⁶ It justified that change by saying it “could open more pathways to law schools, increasing diversity in all its forms, and making it easier for students to pursue joint degrees.”¹²⁷ Other law schools have followed that path, but also report that only a small number of applicants submitted GRE scores rather than LSAT scores.¹²⁸ When the University of Arizona said the move toward the GRE would make the applicant pool more “diverse,” it does not appear that they were primarily referring to racial or socio-economic diversity. No one has published the racial or socio-economic composition of applicants who submit GRE scores, but Professor Aaron Taylor has published compelling evidence in support of

¹²² For further discussion, see MICHAEL SAUDER & WENDY ESPELAND, LAW SCH. ADMISSION COUNCIL, INC., FEAR OF FALLING: THE EFFECTS OF U.S. NEWS & WORLD REPORT RANKINGS ON U.S. LAW SCHOOLS 4 (2007), <https://www.lsac.org/data-research/research/fear-falling-effects-us-news-world-report-rankings-us-law-schools-gr-07-02> (reporting that “the selectivity of the law school accounts for 25% of the overall score” with LSAT score having an overall weight of “12.5% of the overall score”).

¹²³ *Id.* at 12.

¹²⁴ *Id.*; see also Andrew S. Belasco et al., *The Test-Optional Movement at America’s Selective Liberal Arts Colleges: A Boon for Equity or Something Else?*, 37 EDUC. EVALUATION & POL’Y ANALYSIS 206, 208 (2015) (reporting studies that demonstrate “a strong positive correlation between standardized test achievement and socioeconomic status”).

¹²⁵ Aaron Taylor found that “[i]n 2015, respondents with LSAT scores of 155 or below were almost twice as likely to expect more than \$120,000 in law school debt than respondents with higher scores. No such disparities were observed 10 years ago.” Aaron Taylor, *The GRE Is No Diversity Tool*, NAT’L JURIST (June 7, 2016), <http://www.nationaljurist.com/national-jurist-magazine/gre-no-diversity-tool>.

¹²⁶ See *UA Becomes First Law School to Open Admissions to All GRE Test Takers After Study Reveals Test Validity*, U. ARIZ. (Feb. 10, 2016), <https://law.arizona.edu/news/2016/02/ua-becomes-first-law-school-open-admissions-all-gre%C2%AE-test-takers-after-study-reveals>.

¹²⁷ *Id.*

¹²⁸ See Alex Swoyer, *Top Law Schools Accepting GRE Instead of LSAT to Broaden Pool*, WASH. TIMES (Jan. 1, 2018), <https://www.washingtontimes.com/news/2018/jan/1/gre-being-accepted-by-top-law-schools-instead-of-l/> (reporting that the University of Arizona admitted twelve students in 2017 based on GRE scores and that eleven of them enrolled).

the argument that “the GRE is no diversity tool.”¹²⁹

In Harvard Law School’s statement on why it was accepting GRE scores, it said the move would promote diversity “in terms of academic background, country of origin, and financial circumstances.”¹³⁰ The premise of the financial circumstances argument was that students would find it less expensive to apply if they had already taken the GRE and would now not need to take the LSAT. Further, low-income students might find it more convenient to take the GRE than the LSAT because the GRE is offered nearly every day, and the LSAT is offered on a more occasional basis. More realistically, however, students might consider it necessary to take *both* the GRE and the LSAT to see which score puts them in a more advantageous position. Also, given that few students would apply *only* to the dozen or so schools that accept the GRE, it appears that most students would still need to pay to take both tests, even if they have already taken the GRE.

The “academic background” argument would seem to tilt in favor of male applicants. Harvard Law School’s statement also emphasized the increased access that this move would offer to students who focus on “biology, computer science, and engineering.”¹³¹ STEM fields are notoriously male-dominated¹³² and the GRE is math-heavy, so it is hard to see how this move would benefit female applicants as a group. In fact, GRE’s published data reflect that the mean score on the Quantitative Reasoning Section of the test was 150.8 for women (between 38th and 42nd percentile) and 154.9 for men (54th and 58th percentile), while women barely outperformed men, on average, in the Verbal Reasoning Section of the test (150.1 as compared to 149.5).¹³³

The movement toward accepting GRE scores put pressure on the ABA to relax its requirement that alternative tests be validated before being used for admissions purposes because of problems with how that validation would be attained.¹³⁴ The data available for such validity studies was fairly limited

¹²⁹ See Taylor, *supra* note 125.

¹³⁰ See *In Pilot Program, Harvard Law Will Accept GRE for Admission*, HARV. L. TODAY (Mar. 8, 2017), <https://today.law.harvard.edu/gre/>.

¹³¹ *Id.*

¹³² See, e.g., Jessi L. Smith et al., *When Trying Hard Isn’t Natural: Women’s Belonging with and Motivation for Male-Dominated STEM Fields as a Function of Effort Expenditure Concerns*, 39 PERSONALITY & SOC. PSYCHOL. BULL. 131 (2012).

¹³³ See EDUC. TESTING SERV., A SNAPSHOT OF THE INDIVIDUALS WHO TOOK THE GRE GENERAL TEST 12 (July 2013–June 2016), https://www.ets.org/s/gre/pdf/snapshot_test_taker_data_2016.pdf. For percentiles, see EDUC. TESTING SERV., GRE GENERAL TEST INTERPRETIVE DATA 1–2 (2018), https://www.ets.org/s/gre/pdf/gre_guide_table1a.pdf.

¹³⁴ See ABA STANDARDS AND RULES OF PROCEDURE, *supra* note 24, Interpretation 503-1 (“A law school that uses an admission test other than the Law School Admission Test sponsored by the Law School Admission Council shall demonstrate that such other test is a valid and reliable test to assist the school in assessing an applicant’s capability to satisfactorily

since few applicants had taken the GRE before enrolling in law school. Nonetheless, once a few schools had validated the use of the GRE for admissions, it seemed odd to require each additional school to run its own validity study.¹³⁵

In March 2017, the ABA proposed a change to Standard 503 that would permit law schools to use any admission test that was valid and reliable for law school admission.¹³⁶ But then, in November 2017, it changed course and proposed a modification to ABA standards that would eliminate Standard 503's test requirement and modify Standard 501 to give law schools great flexibility to establish admission criteria with or without standardized test scores. Proposed Standard 501(c) would provide:

Among the factors to consider in assessing compliance with this Standard [requiring sound admissions policies and practices] are the academic and admission credentials of the law school's entering students, the academic attrition rate of the law school's students, the bar passage rate of its graduates, and the effectiveness of the law school's academic support program. Compliance with Standard 316 is not alone sufficient to comply with the Standard.¹³⁷

Further, Interpretation 501-1 would generally permit law schools to comply with Standard 501 by not requiring the use of standardized test scores:

Sound admissions policies and practices may include consideration of admission test scores, undergraduate course of study and grade point average, extracurricular activities, work

complete the school's program of legal education.").

¹³⁵ This Article is not going to challenge the validity of the GRE as a testing instrument to determine if students should be admitted to law school, but it is a subject worthy of serious investigation (aside from the GRE's troubling speeded aspect). Because the GRE is fifty percent math, and because some of the math is at a very advanced academic level, it is hard to understand why the score would be a strong predictor of success as a law student or lawyer. This percentage is especially troubling because of the data discussed above showing that men, on average, outperform women on the math section. *See supra* note 133. In view of this, the use of the GRE is likely to lessen gender diversity in law schools.

¹³⁶ *See* Memorandum from Gregory M. Murphy, Council Chairperson, Am. Bar Ass'n & Barry A. Currier, Managing Dir. of Accreditation and Legal Educ., Am. Bar Ass'n, to Interested Persons and Entities (Mar. 24, 2017) (alterations in original omitted), https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/council_reports_and_resolutions/20170324_notice_and_comment_memo_authcheckdam.pdf.

¹³⁷ Memorandum from Maureen A. O'Rourke, Council Chair, Am. Bar Ass'n & Barry A. Currier, Managing Dir. of Accreditation and Legal Educ., Am. Bar Ass'n, to Interested Persons and Entities (Nov. 17, 2017) (alterations in original omitted), https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/council_reports_and_resolutions/20171117_notice_and_comment_authcheckdam.pdf.

2019]

TEST VALIDITY

711

experience, performance in other graduate or professional programs, relevant demonstrated skills, and obstacles overcome. If a law school requires an admission test, it shall publish information regarding which tests are accepted.¹³⁸

The ABA then sought comment on those proposed changes.¹³⁹ It received comments from twenty individuals or organizations, and has not yet decided whether to accept the proposed changes.¹⁴⁰ The commentators tended to support the exclusive use of the LSAT or the possibility of a standardized test being optional for admissions purposes. Although some commentators criticized the way that law schools use (or overuse) the LSAT for admissions purposes, no commentator criticized the structure of the LSAT itself—its content or speed constraint. Groups that supported racial diversity both supported and opposed these proposed revisions.¹⁴¹

While no commentators mentioned the speeded nature of the LSAT as a testing instrument, some commentators discussed the way the LSAT was being used as a cut off score for considering applicants. In its comments to the ABA, The Society of American Law Teachers (SALT) reported the following:

In recent history, it has been noted that the ABA has denied accreditation to schools that admitted students with LSAT scores lower than 143. In fact, it has been noted that the ABA denies

¹³⁸ *Id.*

¹³⁹ *See id.*

¹⁴⁰ *See Notices of Proposed Standards Changes and Responses to Proposed Standards Changed: Notice and Comment*, A.B.A. (June 19, 2018), https://www.americanbar.org/groups/legal_education/resources/notice_and_comment.html. The ABA received comments from

American Indian Law Center; R. Lawrence Dessem; Kevin K. Washburn; Cristina B. Whitman; AccessLex Institute; Seton Hall University School of Law; Minority Network; Interested Admissions Deans and Directors; Educational Testing Service; SALT; CLEO; Robert A. Williams Jr.; LSAC; Bradley J.B. Toben; Clinical Legal Education Association; Deans Miller, Chemerinsky, Rodriguez, and Farnsworth; Daniel B. Rodriguez; and Benjamin Brown.

Id.; *see also* Stephanie Francis Ward, *Legal Education Council Delays Action on Questionnaire Guidance*, A.B.A. J. (Aug. 3, 2018), http://www.abajournal.com/news/article/law_school_guidance_regarding_possible_change_in_entrance_exam_reporting_po/.

¹⁴¹ *Compare* Letter from Cassandra Sneed Ogden, Chief Exec. Officer, Council on Legal Educ. Opportunity (CLEO), to Maureen A. O'Rourke, Council Chair, Am. Bar Ass'n (Apr. 2, 2018), https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/council_reports_and_resolutions/comments/503_cleo.authcheckdam.pdf (opposing elimination of Standard 503), *with* Letter from Christopher P. Chapman, President & Chief Exec. Officer, AccessLex Inst., Members of the Council of the Am. Bar Ass'n Section on Legal Educ. and Admissions to the Bar (Mar. 19, 2018), https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/council_reports_and_resolutions/comments/20180319_comment_s501_s503_acceslex_institute.authcheckdam.pdf (supporting elimination of Standard 503).

accreditation to schools that admit any applicant with an LSAT score of 140. Therefore, the practical effect of these accreditation denials is a de facto “cut-off score” range of 141-143. An even higher cut-off score has been reported among New York schools which reportedly do not accept students with LSATs below 150.

LSAC recently released data establishing an average LSAT score of 143.5 for African-American students. Application of the ABA de-facto standard to this average would mean the automatic disqualification of half of the African-Americans who take the LSAT.¹⁴²

The comments about applicants whose scores are in the bottom half or quartile for test takers raises an implicit validity question. Is this test even valid for test takers who score below 150 (the median score)? Because test takers who score in the bottom half of the applicant pool are likely to have not completed the entire exam to the best of their ability, the exam may have tested their speediness but may not have tested their aptitude for law school (beyond being good at speeded exams). If a standardized test were to be required for admissions purposes, would that test be more useful or valid if it were less speeded so that students in the bottom half of the applicant pool would have a better opportunity to demonstrate their skills and abilities? That question will be pursued further in Part V of this article.

So far, this Article has used the psychometric literature to suggest that the speeded nature of standardized exams may undermine their predictive validity. Are there any *legal* arguments that entities should be required to demonstrate the validity of the speeded nature of a standardized exam? The answer is that there are some legal arguments available from the *employment* context under Title VII in favor of such validity studies in the race and gender context, but those arguments do not technically apply to *admissions tests* to institutions of higher education. In order to find such arguments in the admissions context, one has to turn to the ADA, which applies to disability discrimination, but not to race or gender discrimination. The melding of the arguments available in both contexts provides the strongest legal argument, as discussed in Parts III, IV, and V, below.

¹⁴² Letter from Matthew H. Chairty, Co-President & Davida Finger, Co-President, Soc’y of Am. Law Teachers (SALT), to Members of the Council of the Am. Bar Ass’n Section on Legal Educ. and Admissions to the Bar (Mar. 31, 2018), https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/council_reports_and_resolutions/comments/20180331_comment_s503_salt_authcheckdam.pdf (internal footnotes omitted) (submitted for the April 21, 2018 meeting of the Standards Review Committee).

III. TITLE VII PRINCIPLES

A. *Employment Context*

The courts' foray into the area of standardized testing received impetus from two of the classic Title VII race discrimination cases: *Griggs v. Duke Power Company*¹⁴³ and *Albemarle Paper Company v. Moody*.¹⁴⁴

The story in *Griggs* is an excellent reminder of the ways in which testing rules can be used deliberately to bar racial minorities from certain employment opportunities. Before Title VII went into effect, Duke Power Company only employed African-Americans in labor positions, where the highest paid employee in that department was paid less than the lowest paid employee in other departments.¹⁴⁵ On the date on which Title VII became effective, the employer began to require all potential employees to receive a satisfactory score on two professionally prepared aptitude tests (the Wonderlic Personnel Test and the Bennett Mechanical Comprehension Test) in order to qualify for placement in any department other than the Labor Department.¹⁴⁶

When African-American plaintiffs challenged this new rule under Title VII, the lower courts found that those requirements were permissible under Title VII because they were "applied fairly to whites and Negroes alike."¹⁴⁷ The United States Supreme Court overturned the lower court decisions, finding that hiring rules, such as the testing requirement, which adversely impact a racial group, must "be shown to be related to job performance."¹⁴⁸

In rendering this decision, the statutory language that gave the Supreme Court the most pause was § 703(h) of Title VII.¹⁴⁹ It provides:

Notwithstanding any other provision of this subchapter, it shall not be an unlawful employment practice for an employer . . . to give and to act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex or national origin.¹⁵⁰

This rule was ambiguous because it seemingly protected "professionally developed ability tests," while also prohibiting such tests from being "designed, intended or used" to discriminate on the basis of race.

¹⁴³ 401 U.S. 424 (1970).

¹⁴⁴ 422 U.S. 405 (1975).

¹⁴⁵ *Griggs*, 401 U.S. at 427–28.

¹⁴⁶ *Id.*

¹⁴⁷ *Id.* at 429.

¹⁴⁸ *Id.* at 431.

¹⁴⁹ *Id.* at 433–34.

¹⁵⁰ 42 U.S.C. § 2000e-2(h) (2018).

What would it mean for a professionally-developed test to be “used” to discriminate on the basis of race if there is no evidence that it was “designed” or “intended” to discriminate on the basis of race?

Because the lower courts had already found that Duke Power did not choose the professionally developed test for the purpose of excluding African-Americans, the Supreme Court had to emphasize the “use” prohibition under § 703(h) to invalidate the exams. It found that employers violated the “use” prohibition when they allowed such devices to have a “controlling force unless they are demonstrably a reasonable measure of job performance.”¹⁵¹ Applying that standard, the Court concluded that the professionally developed tests violated § 703(h) because they were adopted “without meaningful study of their relationship to job-performance ability.”¹⁵² Further, because employees who had previously been promoted could not necessarily meet those testing standards, the Court found that the company had made “no showing” that they even furthered the “avowed policy of advancement within the Company.”¹⁵³ Thus, those tests could not be used for hiring or promotion purposes.

Griggs was an extremely important case that presented a novel theory—an employer could not use a professionally-developed test for job selection or advancement purposes that produced a racially disparate impact despite § 703(h), unless that test could be shown to be a valid predictor of job performance. In rendering that decision, the Court gave deference to the Equal Employment Opportunity Commission’s (EEOC) enforcement guidelines, which stated: “[t]he fact that a test was prepared by an individual or organization claiming expertise in test preparation does not, without more, justify its use within the meaning of Title VII.”¹⁵⁴ Employers could no longer pick a test off the shelf that *sounded* valid because of its appealing name, like “the Wonderlic Personnel Test” or “the Bennett Mechanical Comprehension Test.” They had to demonstrate that the test was a valid measure of job performance *at their workplace* if it produced disparate racial impact.

After the *Griggs* decision, companies like Albemarle Paper Company, which were also using the Wonderlic Personnel Test, hired an industrial psychologist to study the “job relatedness” of their testing program.¹⁵⁵ The mere hiring of an industrial psychologist, however, was not sufficient to validate a test under § 703(h). In *Albemarle Paper Co. v. Moody*, both the

¹⁵¹ *Griggs*, 401 U.S. at 436.

¹⁵² *Id.* at 431.

¹⁵³ *Id.* at 432.

¹⁵⁴ *Id.* at 433 n.9 (quoting EEOC Guidelines on Employment Testing Procedures, 29 C.F.R. §§ 1607.1–.18 (2018)).

¹⁵⁵ See *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 410–11 (1975) (studying the validity of the Revised Beta Examination and the Wonderlic Personnel Test for a variety of company positions).

Court of Appeals and the Supreme Court found that the test validation study performed by Albemarle Paper Company's expert could not meet professional test validity standards.¹⁵⁶ The expert did not study every job category, used informal and subjective rankings by supervisors to measure job performance, and did not find statistical correlations in many of the job categories between test scores and job performance, yet concluded that the professionally-developed tests could be used for all jobs within the company.¹⁵⁷ Further, the validation study administered the tests only to job-experienced white workers rather than the typical population for job applicants. EEOC guidelines emphasized the importance of validity studies including minority candidates.¹⁵⁸ Whereas the *Griggs* decision required the hiring of industrial psychologists to validate professionally developed tests for the employer's work force, *Albemarle* required those industrial psychologists to meet rigorous standards for test validity.

In subsequent litigation brought by the United States Department of Justice (DOJ), police and fire departments were routinely found to have violated test validity principles when they used various written examinations for selection for entry-level positions.¹⁵⁹ These principles were also used to invalidate (1) the use of a national teacher examination as a condition for retention of teachers;¹⁶⁰ (2) tests used to hire employees at an electric utility;¹⁶¹ (3) a physical examination used to select firefighters;¹⁶² and (4) written exams used to select police officers and fire fighters.¹⁶³ One rule that emerged from this litigation was that plaintiffs could prevail if they demonstrated that "some other practice, without a similarly undesirable side effect, was available and would have served the defendant's legitimate interest equally well."¹⁶⁴

¹⁵⁶ *Id.* at 430–31.

¹⁵⁷ *Id.* at 431–34.

¹⁵⁸ *Id.* at 435.

¹⁵⁹ See, e.g., *United States v. City of Buffalo*, 457 F. Supp. 612 (W.D.N.Y. 1978); *United States v. City of St. Louis*, 418 F. Supp. 383 (E.D. Mo. 1976).

¹⁶⁰ See *Walston, Jr. v. Cty. Sch. Bd.*, 492 F.2d 919 (4th Cir. 1974) (use of cut-off score for teacher retention was arbitrary and discriminatory).

¹⁶¹ See *Stamps v. Detroit Edison Co.*, 365 F. Supp. 87 (E.D. Mich. 1973), *judgment reversed by EEOC v. Detroit Edison Co.*, 515 F.2d 301 (6th Cir. 1975), *judgment vacated*, *Detroit Edison Co. v. EEOC*, 431 U.S. 951 (1977).

¹⁶² See *Brunet v. City of Columbus*, 642 F. Supp. 1214 (S.D. Ohio 1986), *reversed by Brunet v. City of Columbus*, 1 F.3d 390 (6th Cir. 1993).

¹⁶³ See *Ensley Branch of NAACP v. Seibels*, 616 F.2d 812 (5th Cir. 1980).

¹⁶⁴ *EEOC v. Steamship Clerks Union, Local 1066*, 48 F.3d 594, 602 (1st Cir. 1995); see also *Ernst v. City of Chicago*, 837 F.3d 788, 794 (7th Cir. 2016) (finding that an employee can prevail under Title VII "by proving that the employer has rejected an available alternative job practice that (1) results in a less disparate impact, and (2) serves the employer's legitimate needs" (citing *Ricci v. DeStefano*, 557 U.S. 557, 578 (2009)); *Bradley v. City of Lynn*, 443 F. Supp. 2d 145, 157 (D. Mass. 2006) (quoting *Steamship Clerks Union*, 48 F.3d at 602).

Not wanting to spend the money necessary to develop a valid exam, some employers began to look for shortcuts that would allow them to hire a racially balanced workforce while not spending money on a universally valid exam. Their shortcut was to administer a non-validated exam, and then apply different cut scores for determining if white or black applicants were eligible for consideration for an employment position.¹⁶⁵

In 1991, Congress ended that practice by amending Title VII to provide that it was “an unlawful employment practice . . . to adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment related tests on the basis of race, color, religion, sex, or national origin.”¹⁶⁶ Following the passage of the 1991 Amendment, white job applicants brought a Title VII case challenging the validity of testing practices that sought to develop testing instruments which were universally administered, yet were also chosen because they produced a minimal racially adverse impact.¹⁶⁷ In *Hayden v. County of Nassau*,¹⁶⁸ the Second Circuit upheld those kinds of practices as consistent with Title VII when the hiring entity used the same cut-off scores for all applicants.¹⁶⁹

White plaintiffs did, however, succeed in using the 1991 Amendment to prevent an employer from doing its utmost to reduce the racial impact from a promotion exam, after the test was administered. In *Ricci v. DeStefano*,¹⁷⁰ the plaintiffs challenged the decision of the New Haven Fire Department to refuse to certify the results of a promotion examination due to the potential adverse racial effect of the selection instrument. In sharp contrast to *Griggs* and *Albemarle*, the fire department hired industrial psychologists to design promotional examinations, and was careful to include minority firefighters in their work so that the results “would not unintentionally favor white candidates.”¹⁷¹

Despite these efforts, the top ten candidates who were eligible for immediate promotion to lieutenant were white; the top nine candidates who were eligible for promotion to captain consisted of seven whites and two

¹⁶⁵ See, e.g., *Majeske v. City of Chicago*, 29 F. Supp. 2d 872 (N.D. Ill. 1998) (upholding use of different cut-off scores on the basis of race in the administration of a police officer test for making promotions to detective in 1990).

¹⁶⁶ 42 U.S.C. § 2000e-2(l) (2018).

¹⁶⁷ See *Fioriglio v. State of N.J., Dep’t of Pers.*, 166 F.3d 1205 (3d Cir. 1998) (affirming dismissal of complaint by white applicant).

¹⁶⁸ 180 F.3d 42 (2d Cir. 1999).

¹⁶⁹ *Id.* at 46. But see *Gulino v. N.Y. State Educ. Dep’t*, 460 F.3d 361 (2d Cir. 2006) (upholding validity of public school teacher certification program in challenge by African-American and Latino educators).

¹⁷⁰ 557 U.S. 557, 562–63 (2009).

¹⁷¹ *Id.* at 565.

Hispanics, but no African-Americans.¹⁷² After hearing from numerous witnesses about the validity of the test, the adverse impact of the test results, and possible legal liability, the Civil Service Board (CSB) voted whether to certify the test results. “With one member recused, the CSB deadlocked 2 to 2, resulting in a decision not to certify the results.”¹⁷³

In a five to four decision, the Supreme Court concluded that the white and Hispanic plaintiffs had a valid claim of disparate treatment under Title VII. Job applicants were entitled to expect that the test scores, which they had attained after considerable practice and expense, would be the means by which the city chose promotional candidates, especially when the test had been properly validated.¹⁷⁴

Justice Ginsburg authored a strong dissent in which she took issue with how the majority described the facts. She argued that CSB had ample evidence of validity problems with the examination instrument and inequitable access to test-preparation materials, which could cause CSB to fear a valid disparate impact suit if the test results were used.¹⁷⁵ Justice Ginsburg tells the story of an employer that, after hiring an outside consultant to develop a fair and valid test, is still unsatisfied with the test and wants to do more work to make it both valid and fair. Nonetheless, legal intervention curtailed that process so they had to use the test results from what they considered to be a flawed testing procedure and select candidates for promotion.

Both the majority and dissenting opinions, however, agreed that entities have an obligation under *Griggs* and *Albemarle* to develop valid testing practices, and that plaintiffs could prevail if they demonstrate “that ‘other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer’s legitimate interest.’”¹⁷⁶ Further, courts have developed extensive case law under Title VII to add substance to what it means for a test to be “valid.” Tests are not merely valid because they have some facial validity to what might appear to be the job requirements. Instead, rigorous testing is needed to determine how a test measures the skills and abilities actually needed on the job. It is difficult to develop valid testing instruments; employers or others who are committed to fair testing practices must prioritize such practices.

¹⁷² *Id.* at 566.

¹⁷³ *Id.* at 574.

¹⁷⁴ *Id.* at 593.

¹⁷⁵ *Id.* at 613–18 (Ginsburg, J., dissenting).

¹⁷⁶ *Ricci*, 557 U.S. at 578 (quoting *Albemarle Paper Co. v. Moody*, 422 U.S. 405 (1975)).

B. *Outside the Employment Context*

While employers have been struggling for decades to develop fair testing practices, such efforts have been far more muted outside the employment context. One reason is that Title VII litigation has not put pressure on testing outside the employment context.

Attempts to invalidate tests as racially discriminatory outside the employment context have been largely unsuccessful. Although professional exams, such as the Bar Exam, have a disparate racial impact on test takers' employment prospects, those tests have not been reachable under Title VII. The tests are either found to meet *Griggs* validity principles, or are not covered by Title VII because the plaintiff is not suing the "employer."¹⁷⁷

While arguments have been made that the SAT and ACT discriminate against African-American applicants to college, no court has ever accepted those arguments as demonstrating the violation of any relevant federal statute.¹⁷⁸ Professors William Kidder and Jay Rosner have argued that the SAT could be revised to lessen the racial impact, but acknowledge that "ETS and the College Board are very unlikely to adopt impact reduction techniques in connection with the SAT unless outside pressure is so substantial as to impact the SAT marketplace."¹⁷⁹ Instead of persuading ETS or the College Board to modify the ACT or SAT, some colleges have chosen to make the ACT/SAT optional by joining a consortium called the National Center for Fair and Open Testing ("FairTest").¹⁸⁰

Standardized exams outside the employment context, therefore, continue to produce a huge racially disparate impact, but no legal tools exist under Title VII to challenge that impact. Voluntary decisions to make the ACT or SAT optional may have helped lessen some of the disparate impact of those exams,¹⁸¹ but no college or university has requested that the speeded aspect of the ACT or SAT be modified to reduce their racial disparate impact, even though that aspect of the exam may be contributing to the racially disparate impact.

¹⁷⁷ For further discussion, see Linda J. Strassle, Note, *Minimum Competency Testing of Teachers for Certification: Due Process, Equal Protection and Title VII Implications*, 70 CORNELL L. REV. 494 (1985).

¹⁷⁸ See, e.g., Kendra Johnson, *Racially Bias SAT I/ACT Blocks College Access: Is It Constitutional for College Officials to Condition Admission on a Racially Bias Assessment*, 33 U. BALT. L.F. 2 (2003); Kidder & Rosner, *How the SAT Creates "Built-In Headwinds"*, *supra* note 35.

¹⁷⁹ Kidder & Rosner, *How the SAT Creates "Built-In Headwinds"*, *supra* note 35, at 211.

¹⁸⁰ See *More than 1000 Accredited Colleges and Universities that Do Not Use ACT/SAT Scores to Admit Substantial Numbers of Students into Bachelor-Degree Programs*, FAIRTEST, <https://www.fairtest.org/university/optional> (last visited Oct. 27, 2018).

¹⁸¹ See Belasco et al., *supra* note 124, at 214 (reporting conflicting evidence on whether test-optional practice increases diversity).

IV. ADA PRINCIPLES

Testing issues have been resolved under a quite different framework under the ADA than Title VII. They are governed by the following statutory language found in the ADA:

Any person that offers examination or courses related to applications, licensing, certification, or credentialing for secondary or post-secondary education, professional, or trade purposes shall offer such examinations or courses in a place and manner accessible to persons with disabilities or offer alternative accessible arrangements for such individuals.¹⁸²

On its face, § 12189 clearly applies to exams for admissions, such as the LSAT, and exams for professional certification, such as those required by the medical or legal profession.¹⁸³ Unlike the Title VII context, there is no question that these admission and certification exams are covered by the applicable rule.

In the ADA context, the big hurdle facing individuals who want to challenge tests is whether the plaintiffs are covered by the clause requiring them to be a “person with a disability” in order to take advantage of these protections. If plaintiffs satisfy that requirement, then courts proceed to apply the provision itself. By contrast, under Title VII, an individual does not have to belong to a certain subcategory to attain statutory coverage: men and women are covered; blacks and whites are covered.¹⁸⁴

In order to satisfy the “person with a disability” requirement under the ADA under § 12189, plaintiffs typically argue that they have a learning disability that causes them to read more slowly than the non-disabled population. Because most standardized tests are speeded in the sense that a test taker needs to move at a good pace to complete all the test questions, they argue that their disability precludes them from demonstrating that they have the knowledge, abilities, and aptitudes that the test purports to measure.¹⁸⁵

¹⁸² 42 U.S.C. § 12189 (2018).

¹⁸³ Title III of the ADA only covers various kinds of “public accommodations” for most of its rules. *See id.* § 12181(7) (listing the types of “public accommodations”). But § 12189 was deliberately written to cover “any person” so that it could cover entities like professional organizations and testing companies, which are not otherwise covered by Title III of the ADA when they administer examinations. *See id.* § 12189.

¹⁸⁴ *See McDonald v. Santa Fe Trail Transp. Co.*, 427 U.S. 273, 280 (1976) (finding that Title VII was intended to also cover white men and white women).

¹⁸⁵ The regulations provide that any entity offering a covered examination must assure that

[t]he examination is selected and administered so as to best ensure that, when the examination is administered to an individual with a disability that impairs sensory, manual, or speaking skills, the examination results accurately reflect the individual’s aptitude or achievement level or

The wording of § 12189 appears to provide defendants with two options when plaintiffs argue that the test precludes them from demonstrating the skills that the examination purports to measure. Defendants could offer to change the test itself so that is accessible for all test takers by changing its “place and manner.” Or, they could offer “alternative arrangements” for plaintiffs so that the test does seek to measure their actual knowledge, abilities, and aptitudes rather than their disabilities.

So far, the case law has exclusively proceeded under the second option under the assumption that extended time is an “alternative arrangement” that makes an exam “accessible” for the purposes of § 12189. Plaintiffs have used litigation to seek extended time as an “alternative accessible arrangement” under § 12189. The litigation has largely revolved around whether plaintiff qualifies as a “person with a disability” rather than on the issue of whether extended time is a genuine “alternative accessible arrangement.” If courts find that the plaintiff is a person with a disability, then they have typically awarded extended time to plaintiff as a remedy. No one has made the argument that extended time for a subset of test takers with disabilities cannot actually satisfy the requirements of § 12189 when a non-speeded version of the exam for all test takers would actually be the most valid testing instrument. That argument will be developed in Parts IV and V.

The leading case involving the remedy of extended time was decided by Sonia Sotomayor when she was a circuit court judge sitting by designation as a district court judge in *Bartlett v. New York State Board of Law Examiners*.¹⁸⁶ Although this case was decided in 2001, it is probably still one of the most significant ADA cases on the importance of accommodating individuals with disabilities who seek to take professional exams. The significance of this case is that then-Judge Sotomayor understood that a highly intelligent person, like Marilyn Bartlett, who has a slow processing speed in reading, can be a “person with a disability” and thus entitled to relief under § 12189.

Marilyn Bartlett sued the New York State Board of Law Examiners after they denied her request for extended-time accommodations. Her experience is an excellent example of how artificial time constraints can preclude an individual from demonstrating her knowledge and aptitude. Despite having a Ph.D. in English, Marilyn Bartlett consistently scored very

whatever other factor the examination purports to measure, rather than reflecting the individual’s impaired sensory, manual, or speaking skills (except where those skills are the factors that the examination purports to measure).

28 C.F.R. § 36.309(b)(i) (2018).

¹⁸⁶ No. 93 Civ. 4986(SS), 2001 WL 930792 (S.D.N.Y. Aug. 15, 2001).

poorly on speeded standardized exams due to her profound dyslexia. When presented with speeded conditions, she could “read at 195 words per minute” but could only attain a reading comprehension rate at the 16th percentile as compared to college freshmen.¹⁸⁷ By contrast, if she was given a non-speeded reading test, she slowed her reading down to 156 words per minute and achieved a reading comprehension score in the 98th percentile.¹⁸⁸ Her “reading ease level”—the speed at which she attained excellent comprehension—was below the fourth grade level. But, unlike the typical fourth grader, she could attain reading comprehension levels in the “98th percentile . . . when compared to college freshmen.”¹⁸⁹ Thus, she had excellent reading comprehension *if she was given enough time to read the material*.

The New York State Board of Law Examiners, however, concluded that she did not have a disability covered by the ADA and, therefore, was not entitled to relief under § 12189. After years of litigation, then-Judge Sotomayor found in her favor and ordered the defendant to allow her to take the exam with double the normally allotted time, spaced out over four days.¹⁹⁰ As is typical in this kind of litigation, no one asked why it was important for non-disabled candidates to be required to take the exam in a four-hour period over two days. No one asked whether the speeded exam was a more valid predictor of whether one had the skills and abilities to practice law than the non-speeded version of the exam.

Many ADA plaintiffs in testing cases, however, do not even get to the relief stage because they are found not to be a “person with a disability” as required by § 12189. A good example of that problem can be found in *Bibber v. National Board of Osteopathic Medical Examiner, Inc.*¹⁹¹ In *Bibber*, the plaintiff, Bernadette Bibber, was deaf and was diagnosed as having dyslexia. She was denied extended time when she applied to take the professional exam to be licensed in osteopathic medicine because the testing entity considered her not to have a disability that warranted accommodation.¹⁹²

Her story is a typical example of the adverse impact that speeded exams can have on someone’s educational and career opportunities, despite earlier success in an accommodated setting. Bibber was able to attain extended time for the PSAT and SAT, and was admitted to Colby College, a highly selective university.¹⁹³ In order to receive accommodations on the MCAT,

¹⁸⁷ *Id.* at *7.

¹⁸⁸ *Id.*

¹⁸⁹ *Id.*

¹⁹⁰ *Id.* at *51.

¹⁹¹ No. 15-4987, 2016 WL 1404157 (E.D. Pa. Apr. 11, 2016).

¹⁹² *Id.* at *11.

¹⁹³ *Id.* at *2.

she would have needed to be re-evaluated by a qualified psychologist. Because she could not afford to be re-examined, she took the MCAT without accommodations. Her merely average scores on the MCAT were likely a factor in her inability to be admitted to any medical schools at that time. Some years later, she re-took the MCAT, again without accommodations, and re-applied to medical school. This time, she was admitted to the Rowan School of Osteopathic Medicine.¹⁹⁴

As a medical school student, Bibber needed to take an Osteopathic Licensing Examination. Due to her longstanding dyslexia, she wanted to receive extended time on the test and, once again, needed a new evaluation to make that request. Because she was now a student, she was able to obtain a discounted student rate for a new evaluation, thereby reducing the cost. This evaluation, as well as her self-report, again concluded that she had dyslexia but, like Marilyn Bartlett, had strong reading comprehension when provided adequate time.¹⁹⁵

Using the new evaluation, Bibber applied for extended-time accommodations on the Comprehensive Osteopathic Medical Licensing Examination. Her request was denied. Bibber challenged that denial in court and lost. The court concluded that Bibber's reading and processing abilities were "average" when compared to the general population.¹⁹⁶ She was therefore considered not to be a person with a disability for the purposes of § 12189 and was not entitled to any relief.

Bernadette Bibber is not alone in being found to be non-disabled under the "average" rule, which focuses on overall comprehension rather than reading speed. Recently, Elizabeth Black was also denied testing accommodations under that rule. She sued the National Board of Medical

¹⁹⁴ *Id.* at *3.

¹⁹⁵ *Id.*

¹⁹⁶ This "average" problem stems from a particular ADA regulation that contains the language "as compared to most people." It states:

(1) Rules of construction. The following rules of construction apply when determining whether an impairment substantially limits an individual in a major life activity.

....

(v) An impairment is a disability within the meaning of this part if it substantially limits the ability of an individual to perform a major life activity as compared to most people in the general population. An impairment does not need to prevent, or significantly or severely restrict, the individual from performing a major life activity in order to be considered substantially limiting. Nonetheless, not every impairment will constitute a disability within the meaning of this section.

28 C.F.R. § 36.105(d)(1)(v) (2018). The argument is made by some testing entities, as reflected in the *Bibber* case, that test takers with high cognitive aptitude and strong reading comprehension who read slowly are not limited "as compared to [most people in] the general population," and are therefore not persons with disabilities. 2016 WL 1404157, at *7.

Examiners after she was denied extended time on the Step 1 Medical Examination.¹⁹⁷ Rather than considering evidence of her slow rate of reading, the district court judge found that “Black’s history of superlative academic performance refutes the claim that ADHD substantially limits Black’s ability to learn, read, remember, or concentrate in comparison to the average person.”¹⁹⁸ Because she was unable to pass the Step 1 Medical Examination with regular time, this decision effectively ended her medical school career after four years of college, four semesters of pre-med courses at the University of Pennsylvania, and two years of medical school.¹⁹⁹

These cases all revolved around one question: whether the plaintiff is sufficiently disabled so as to be entitled to relief under § 12189? Assuming the plaintiff does have a disability, no one asked whether the time limits were necessary to maintain the validity of the testing instrument. No one sought to devise a remedy that would be available to *all* test takers and not just those who hire a lawyer to request extended time under the “alternative arrangements” language in § 12189.

In Part II of this Article, we saw how difficult it is to determine the precise amount of extended time that an individual student should be allocated so that a test fairly measures the skills and abilities it is designed to measure. Part IV has shown that another problem exists: How impaired does a person need to be to qualify for a remedy under § 12189? Should we focus on a person’s reading rate or a person’s overall reading comprehension to determine if they have a disability? How slow is too slow, so that a person is considered impaired?

A strength of the universal design solution, under which all test takers have an opportunity to take a non-speeded exam, is that we get to avoid both of these thorny issues. As will be discussed in Part V, the ADA can be used to impose the universal design remedy under § 12189 because the non-speeded version is likely to be the most valid testing instrument for all candidates. The basis of that legal argument will be sketched out in Part V by blending Title VII test validity principles into § 12189 of the ADA.

V. A TITLE VII/ADA SYNTHESIS

A. *Importing Title VII Validity Principles into Section 12189*

Thus far, we have seen that speeded exams have a disparate impact on individuals with disabilities, yet little attention has been devoted to justifying those speeded conditions. In the Title VII context, when exams produce a

¹⁹⁷ See *Black v. Nat’l Bd. of Med. Exam’rs*, 281 F. Supp. 3d 1247, 1248 (M.D. Fla. 2017).

¹⁹⁸ *Id.* at 1249–50.

¹⁹⁹ *Id.* at 1250–51.

disparate impact on the basis of race, courts have repeatedly required the test developer to validate the examination instrument and, if available, use less-impactful selection criteria to select their workforce.²⁰⁰

Neither of these inquiries have occurred under § 12189. Despite the well-known disparate impact of timed testing instruments on individuals with disabilities, no one has asked test developers to justify the speeded conditions as a necessary aspect of the validity of these testing instruments. Moreover, no one has asked whether a non-speeded instrument might equally (or better) serve the testing entity's need to distinguish among test takers on the basis of their skills and abilities. Those test validity requirements should be imported into ADA § 12189 by observing the inadequacy of the traditional extended time remedy. Because extended time is an imperfect remedy, testing entities should be required to justify the use of time limits for all candidates and explore the benefits of a non-speeded exam with respect to both disparate impact and test validity. If a non-speeded exam is equally (or more) valid, then § 12189, like Title VII, should require that approach in order to avoid the adverse disparate impact of the testing instrument.

This non-speeded approach would create a universal design solution rather than an accommodation solution to the disparate impact caused by speed constraints. It would be consistent with the longstanding doctrine under Title VII where employers are expected to devise a valid testing instrument used by *everyone* rather than different testing requirements for different groups.

The universal design, non-speeded approach is consistent with the regulations that the DOJ has promulgated under § 12189. These regulations require that an examination be “selected and administered so as to *best ensure* that . . . the examination results accurately reflect the individual's aptitude or achievement level or whatever other factor the examination purports to measure.”²⁰¹ The “best ensure” language supports the requirement that test developers have the burden of proof to demonstrate that they have chosen the most valid testing instrument to assess an individual's skills and abilities. If they insist on using a speeded exam, which has a disparate impact on test takers with disabilities, then they would have the obligation to demonstrate that a speeded testing instrument is more valid than a non-speeded testing instrument to evaluate the skills and abilities they seek to assess. Although this regulation is the first clause of the DOJ's testing regulations, it has not been used to require test developers to validate the speeded nature of exams. They have not been required to justify why a

²⁰⁰ See *supra* Part III.A.

²⁰¹ 28 C.F.R. § 36.309(b)(1)(i) (2018) (emphasis added).

speeded exam is necessary at all, and why a time-extended version of the exam for test takers with disabilities is more valid than a non-speeded exam for all test takers.

From a purposive perspective, the universal design solution is a superior approach to the current “extended time” approach because all candidates are treated the same. Under the ADA (and other disability statutes), integrated solutions are considered superior to segregated solutions. For example, under the Individuals with Disabilities Education Act,²⁰² we insist that children be kept in the general education classroom rather than be pulled out for instruction in a disability-segregated environment.²⁰³ Similarly, under the ADA, the Supreme Court ruled in *Olmstead* that integrated rather than segregated solutions should take priority.²⁰⁴ Yet, under existing ADA law interpreting the testing rules, entities have been allowed to jump to extended time accommodations for disabled test takers without justifying why speeded conditions are required for any test takers.

Under the ADA’s integration principle,²⁰⁵ it makes sense to integrate through a universal design solution rather than to segregate test takers by offering extended time to some but not all.²⁰⁶ Ironically, one of the arguments against offering extended time to individuals with disabilities is that *all* students might benefit from extended time—thus, it seems unfair from that vantage point to only give it to students with disabilities. A universal design solution avoids that problem by providing ample time to *everyone*. All students are offered an opportunity to take as much time as they need to demonstrate their knowledge and abilities.

²⁰² Individuals with Disabilities Education Act, 20 U.S.C. §§ 1400–1445 (2018).

²⁰³ See, e.g., *T.R. v. Kingwood Twp. Bd. of Educ.*, 205 F.3d 572, 578–79 (3d Cir. 2000) (defining the “least restrictive environment” as one in which the school district “to the greatest extent possible, satisfactorily educates disabled children together with children who are not disabled, in the same school the disabled child would attend if the child were not disabled” (quoting *Carlisle Area Sch. v. Scott P.*, 62 F.3d 520, 535 (3d Cir. 1995)).

²⁰⁴ See *Olmstead v. L.C. ex rel. Zimring*, 527 U.S. 581, 600 (1999) (recognizing that “unjustified institutional isolation of persons with disabilities is a form of discrimination”).

²⁰⁵ “Universal design for learning (UDL) is a framework to improve and optimize teaching and learning for all people based on scientific insights into how humans learn.” *About Universal Design for Learning*, CAST, <http://www.cast.org/our-work/about-udl.html#.Wvsjzy-ZM9w> (last visited Feb. 13, 2019).

²⁰⁶ Congress makes reference to the use of universal design principles in the K–12 education context through its partnership grant program. See 20 U.S.C. § 1022a(d)(1)(A)(i)(II) (making funds available for early childhood education programs that include “strategies consistent with the principles of universal design for learning”). Congress also repeatedly references universal design for learning in the Higher Education Opportunity Act of 2008, Pub. L. No. 110-315, 122 Stat. 3078 (codified as amended at 20 U.S.C. § 1001 (2008)). See, e.g., § 202 (teacher quality partnership grants); § 204 (accountability and evaluation); § 205 (accountability for programs that prepare teachers); § 231 (enhancing teacher education); § 251 (teach to reach grants).

This approach has two further advantages: it makes a disability diagnosis less dichotomous while also reducing the socioeconomic bias that can seep into disability determinations. The *Bartlett*, *Bibber*, and *Black* courts had to struggle with the question of “how slow is too slow to be considered ‘disabled’?” Arbitrary lines get drawn between individuals whose reading rate is at the 2nd versus 16th versus 50th percentile in determining whether they have a disability and qualify for extended time. Yet, all those individuals might not have enough time to adequately answer the questions on an exam. If everyone is given extended time, we avoid the fairness question of whether to give extended time only to individuals in the 2nd percentile and not to individuals in the 16th percentile.

The socioeconomic bias of disability law is also evident from these stories. Had *Bibber* been able to afford psychological testing when she applied to take the MCAT, she might have had a stronger record of testing accommodations, as well as more success in gaining admission to medical school. Her inability to be able to afford psychological testing when she took the MCAT cascaded into making it harder for her to attain an accommodation on the licensing exam. Similarly, *Black* had obtained inexpensive testing at her university, which did not satisfy the licensing professionals or the district court. Had she been able to afford more sophisticated testing professionals, she might have met the professional licensing requirements for accommodation. More importantly, if *all* applicants are provided an appropriate amount of time to take these exams, then few applicants would need to spend thousands of dollars on psychological testing.

This approach also helps deal with some of the socioeconomic bias that is well recognized in the testing world because socioeconomic status is a strong predictor of standardized test scores.²⁰⁷ While there are many explanations for this phenomenon, one likely factor is that it takes extensive preparation to develop the skills needed to work quickly on an exam like the LSAT. Testing candidates who do not have to juggle full-time jobs and childcare responsibilities are likely to have more time to develop speed through practice. While a non-speeded exam will not miraculously eliminate socioeconomic advantage on standardized exams, it may help mitigate some of that advantage.

The current approach, therefore, produces unintended consequences of falsely viewing disability as a dichotomous variable and as magnifying socioeconomic inequities. If we changed the default rules for everyone by eliminating speeded conditions, then the playing field could be further

²⁰⁷ See COLL. BD., 2012 COLLEGE-BOUND SENIORS, *supra* note 13, at 4 tbl.11 (demonstrating linear correlation between socioeconomic status and scores on SAT).

2019]

TEST VALIDITY

727

leveled for everyone.

B. *Application of the Test Validity Approach to the LSAT*

1. Existing Validity Data

It is well known that a speeded exam, such as the LSAT, will produce a disparate impact on individuals with disabilities. The standard response has been to provide extended time to such test takers rather than to modify the test for the entire testing population. But what if we required LSAC to validate the speed constraints of the exam?

One could hope that many of the critics of the new flexibility provided by the ABA rule change would welcome the suggestion that we assess whether the speeded nature of the LSAT is necessary. For example, as early as 2011, Jeffrey Zavrotny, Director of Admissions at the University of Baltimore School of Law, expressed opposition to law schools being permitted to admit students without having a standardized test score.²⁰⁸ He said that admissions would be more difficult with less information. But, if the LSAT were offered as a non-speeded exam and offered more information for the bottom half of the applicant pool, law schools like the University of Baltimore would have *more* information rather than less information.

Based on some available, yet limited, validity data, it appears that LSAC may do an especially poor job of predicting law school performance for those who score in the bottom half of the pool (and likely engage in significant guessing behavior). As shown in Part II, the LSAT is a highly speeded exam, with even highly qualified test takers, such as those attending the Moritz College of Law, reporting that they were not able to complete the exam. This kind of information makes one wonder whether the test is a valid predictor of performance, especially for those test takers who guess on large portions of the exam and score in the bottom half of test takers. LSAC has done some evaluation of what is called the “differential prediction” value of the LSAT for various racial groups, but it has never studied test takers based on their scoring band.

The differential prediction study of the performance of racial minorities yields some useful data. A recent LSAT research report “summariz[ing] the results of the 2012–2015 LSAT Correlation Studies . . . based on the 2011–2014 entering law school classes” concluded that the LSAT over-predicts the

²⁰⁸ See Danny Jacobs, *ABA Rethinking LSAT Requirement: Any Change to Test Is Still Years Away*, DETROIT LEGAL NEWS (Feb. 18, 2011), <http://www.legalnews.com/detroit/874133>. Zavrotny is not the only person to express concern about law schools admitting students without standardized test scores. See also Elura Nanos, *Lawyers Are About to Get a Whole Lot Dumber*, LAW & CRIME (Oct. 12, 2017), <https://lawandcrime.com/uncategorized/its-about-to-get-a-lot-easier-to-become-a-lawyer>.

first-year grades of African-American applicants by 2.30 points, of Asians by 2.12 points, and of Hispanics by 1.30 points.²⁰⁹ This “overprediction” is considered acceptable because it supposedly demonstrates that the LSAT is not harming the admissions process for the various affected racial subgroups in comparison to white applicants.²¹⁰

The data, however, are subject to many other interpretations. First, it is important to note that individuals are only a part of this study if they are admitted to and then attend law school. Due to the cut-off data information previously reported, we know that many African-American applicants never have an opportunity to attend law school to learn how they would have performed despite their LSAT score.²¹¹ Further, this “overpredicting” information is part of another statistic—that the correlation coefficient for the LSAT as a predictive tool for African-American applicants is lower than it is for white applicants (0.28 compared to 0.34), and the standard deviation is larger (0.18 compared with 0.10).²¹² The predictive value of a testing instrument is understood to be the square of the correlation coefficient. Thus, the LSAT is able to predict 7.8% of the variance in first-year grades for African-American students as compared with 11.56% of the variance for white students.

Why is the LSAT such a poor predictor of first-year grades for African-American applicants? There are many possible explanations, including problems with the way law schools assign first-year grades and an institutional culture that might not promote learning for all students.²¹³ Another reason may be that the LSAT is an especially poor predictive instrument for those who score below 150 on the exam and likely did not complete all the exam questions (without randomly guessing). The study reported that the mean score on the LSAT for African-American test takers

²⁰⁹ NAZIA RAHMAN & TAMMY J. TRIERWEILER, LAW SCH. ADMISSION COUNCIL, INC., ANALYSIS OF DIFFERENTIAL PREDICTION OF LAW SCHOOL PERFORMANCE BY RACE/ETHNICITY BASED ON 2011–2014 ENTERING LAW SCHOOL CLASSES 14 (2017), <https://www.lsac.org/data-research/research/analysis-differential-prediction-law-school-performance-raceethnicity-based>.

²¹⁰ *Id.* at 3.

²¹¹ The 149 mean data for African-American students is for students *admitted* to law school. *Id.* at 3, 9. LSAC does not publish data on the scores of all test takers by race.

²¹² *Id.* at 13 tbl.6.

²¹³ See Susan Sturm & Lani Guinier, *The Law School Matrix: Reforming Legal Education in a Culture of Competition and Conformity*, 60 VAND. L. REV. 515, 516 (2007) (arguing that the traditional design of law school curriculum “contributes to law student disengagement, particularly for women and people of color”). The traditional first-year law school exam is a timed instrument that may mimic the features of the LSAT. Because the number of words that a student can write in the assigned time is often thought to correlate with the student’s grade, these tests arguably place a very high premium on speediness. This Article does *not* presume that traditional first-year exams are valid or appropriate. That issue is beyond the scope of this Article.

who enrolled in law school was 149.89.²¹⁴ Thus, one should expect that about half of the African-Americans in the study scored below a 150 on the LSAT.²¹⁵ Their lower overall LSAT scores may explain the weak correlation data for this group. In other words, the LSAT may not be a useful predictive instrument for the half of this pool that scored below a 150. LSAC has not conducted a differential validity study to demonstrate if the LSAT is an equally valid predictor of first-year grades for those who score above 150 and those who score below 150.

2. Arguments Against Validity of LSAT

Without a proper validity study regarding the speeded nature of the LSAT, one has to speculate about the validity of the LSAT based on logical arguments and the limited available data. First, let us consider a logical argument to show the problematic nature of using such a speeded exam to predict performance in law school. Let us assume we are giving 100 applicants a 20-question, 30-minute test to measure their reading comprehension and analytical reasoning skills. We curve the test. The median grade is 10 (of 20) correct answers. We tell students to work on the test for 30 consecutive minutes and then to stop working. They should not guess at answers they did not get to. Students G and H receive a raw score of ten on the test and receive a standardized score of 100 (50th percentile). Student G finished every question (20) and missed half (10) of them in order to receive a raw score of 10. Student H proceeded very slowly and only finished 11 of 20 questions. Student H answered 10 of 11 questions correct, but left 9 questions blank. Further, let us assume we used a 10 as a cutoff score for admissions purposes. We provisionally admit no one with a score of 9 or below. We would not admit either of those students although their profile are really very different.

Now, let us assume we only have one spot left in the class and we want to admit one of those two students. It would be helpful to know how these students would have done with additional time. Would Student G have checked her work and found some of her errors? Would Student H have finished and continued with such a high percentage of right answers? More importantly, the validity of the test is likely lessened by the lack of information available about students who do not complete the exam. It seems likely that we would know more about the comparative qualifications of students G and H if they had been given more time to complete the exam.

If the test were an algebra test, it is unlikely that we would accept the possibility that the time limitation enhanced the validity of the testing

²¹⁴ RAHMAN & TRIERWEILER, *supra* note 209, at 9 tbl.5.

²¹⁵ The study reported “mean” not “median,” but these figures are probably close since 150 is in the middle of the 120 to 180 score range.

instrument.²¹⁶ If we want to know whether someone knows 20 discrete elements of algebra, we would want them to answer all 20 questions. We would not presume that a student who answers 11 of 20 algebra questions (and gets 10 correct) does not have the skills to correctly answer the remaining 9 questions. Nor would we think it likely that the student who only answered 11 questions (and got 10 correct) has the same knowledge of algebra as someone who completed all 20 questions but only answered 10 correctly. As applied to the LSAT, test developers would have to demonstrate that the current speeded LSAT is a *better* predictor of a test taker's success in law school than a less-speeded version of the LSAT. In other words, they would have to prove the counterintuitive notion that we have a better sense of Student H's abilities by giving him a test he cannot even complete in the allotted time than by giving him a test he could complete.

Second, let us also consider the static nature of the LSAT despite the changing landscape of legal education. The LSAT has remained essentially unchanged since 1993.²¹⁷ Yet, legal education has changed markedly since 1993, with increased emphasis on writing, formative experiences, and clinical education. The LSAT merely purports to predict the most traditional aspect of legal education—performance on first-year exams in large classes with disproportionately speeded testing instruments. Its speeded structure has not kept pace of innovation in legal education where many professors are seeking other methods of evaluating students. In 2004, when most law schools still used an in-class timed instrument for all first-year courses, William Henderson found that the LSAT did a poor job in predicting grades in upper-year classes using take-home exams or papers as the primary means of assessment.²¹⁸

Although LSAC has published no studies reporting the validity of its testing instrument to predict grades in the first year of law school on a *non-speeded* testing instrument, the author has gathered some data on that subject. In an empirical study with other researchers, the author found that the LSAT was *not* a significant factor in predicting grades in her first-year Constitutional Law class on a non-speeded testing instrument (after controlling for various other factors such as first semester grades).²¹⁹ By

²¹⁶ It is now well accepted that math tests should not be timed because timed tests cause math anxiety “leading to low achievement, math avoidance, and negative experiences of math throughout life.” Jo Boaler, *Research Suggests Timed Tests Cause Math Anxiety*, 20 TEACHING CHILD. MATHEMATICS 469, 469 (2014).

²¹⁷ Interview with Kellye Testy, President and Chief Exec. Officer, Law Sch. Admission Council, Inc., in Newton, Pa. (June 20, 2018).

²¹⁸ See Henderson, *supra* note 12, at 10–11.

²¹⁹ Ruth Colker et al., *Formative Assessments: A Law School Case Study*, 94 U. DET. MERCY L. REV. 387, 404 tbl.6 (2017).

contrast, the LSAT *was* a significant factor in predicting grades from speeded exams in these students' other first-year classes after controlling for the same factors.²²⁰ As more professors move toward exams that are not speeded in the first year of law school,²²¹ one might expect the LSAT to become a less-valid predictor of first-year grades. Nonetheless, LSAC has not sought to control for exam type in determining the validity of the LSAT to predict first-year grades.²²²

3. Response to Arguments in Favor of a Speeded Exam

When the author suggested to LSAC test developers that they consider moving to a non-speeded testing instrument, they expressed concern that such a change would cause the testing instrument to have to contain fewer questions, which, in turn, would lessen the validity of the testing instrument.²²³ They presumed that an exam with fewer questions would be less valid because the testing entity would have less information with which to work in distinguishing test takers. Logically, it is obvious that it is easier to make distinctions among the performance of members of a group if they each take 100 questions, for example, rather than 50 questions.

But the argument that more test questions makes an exam more valid presupposes that most test takers complete each test item. If a test has 100 items, but the average test taker only completes 50 of them, then the length of the test is misleading. The length of the test is only helpful for those who complete all 100 test items. For the other candidates, a part of their score is based on how well they randomly guess. Thus, it is likely that a testing instrument with more test questions would be more valid for distinguishing among some candidates (those who finish) and less valid for distinguishing among other candidates (those who do not finish).

From a law school admissions perspective, we might especially want information about students who are in the bottom half of the curve and are at risk of failing the bar exam. A shorter, non-speeded exam might actually be a better instrument for distinguishing between those candidates. In other words, our current method of testing may make the LSAT, for example, especially useful for the top law schools that have many candidates with

²²⁰ *Id.* at 410 tbl.12.

²²¹ While the author is not aware of a study examining how many professors give non-speeded exams in first-year courses, many professors have told the author that they have decided to move in that direction.

²²² Henderson has published a study examining the validity of the LSAT to predict grades in classes with papers, take-home exams, and in-class exams. He found that the LSAT was a relatively weak predictor of grades on take-home exams and papers. *See* Henderson, *supra* note 12, at 1023–24.

²²³ The author met with LSAC staff on June 20, 2018, and discussed many of the ideas presented in this Article.

scores between 170 and 180, but may make it less useful for the law schools that have many candidates with scores below 150. For the students who score below a 150, it is misleading to describe the LSAT as an exam with 100 questions. For them, it may be an exam with 60 or 70 questions, with the rest of their score being determined by their luck at rapid guessing behavior. Ironically, while the LSAT tests applicants' abilities to identify assumptions behind an argument, the LSAT test developers are making an unproven assumption about their own testing instrument—that it would be less valid if it contained fewer questions.

It is true that the LSAT may be less able to distinguish among the students at the very top of the curve if the exam were not speeded. Instead of distinguishing between a 120 and a 180 using 60 points, the test developers might only be able to distinguish between a 140 and a 160 using 20 points. Rather than 10 points being available to distinguish between the 98th and 99th percentile, only 2 points might be available to distinguish those students. That change, however, might constitute an *improvement* in the law school admissions process, because it would force the top-ranked schools to give less weight to small differences in LSAT scores. All the students who currently receive between a 170 and 180 might receive a 160 on the new scale because many more candidates are likely to get near-perfect scores with extended time for everyone. And, at the bottom of the scale, the difference between a 140 and 142 might be meaningful because the student who earned only a 140 was not able to demonstrate good reading comprehension, irrespective of the time available, while the student who earned a 142 might be a significantly better reader. They are not just separated because of the luck involved in their guessing behavior for questions they did not reach, because all test takers were given an opportunity to attempt every question.

Moreover, less precision on standardized test scores might force admissions offices to give more weight to other factors, such as undergraduate grades, work experience, and personal essays. In other words, the current scale creates an aura of false precision by suggesting that an applicant who scored a 172 is meaningfully different for admissions purposes from an applicant who scored a 180. Both applicants are likely to have the intellectual capacity to do well at any law school; whether they do well will likely be influenced by other factors such as their effort, diligence, and growth mindset.

Despite the absence of studies validating the speeded element of the LSAT, some people justify the speeded nature of the LSAT by saying lawyers generally charge by the hour.²²⁴ Thus, the argument is that it is not

²²⁴ See, e.g., Michelle Kim Hall, *3 Ways the LSAT Prepares Students for Law School*, U.S. NEWS & WORLD REP. (July 10, 2017), <https://www.usnews.com/education/blogs/law-admissions-lowdown/articles/2017-07-10/3-ways-the-lsat-prepares-students-for-law-school>

fair to the legal profession to have a slow lawyer charge by the hour. Of course, not all lawyers charge by the hour. Some lawyers charge by the work product. People can also work in policy or government, where their law license may be appreciated but not required, and where hourly billing does not occur. Further, slow lawyers can adjust their rate to be consistent with market principles, if they choose, or they can agree to cap their rate. Finally, few lawyers would say that the best way to practice law is to race through one's work as quickly as possible, as one might do on a standardized exam. Such a method would likely cause incomplete research or sloppy errors. In fact, legal writing faculty sometimes teach students *not* to prematurely stop their research when they think they have the correct answer through a few citations. Even if a few citations, which are easily located, look helpful, one still needs to go through a pile of other cases to construct the best legal argument. Faster is not necessarily better. Thus, a senior partner in a law firm would rarely ask a junior associate to prepare a polished written product in a few hours; it is likely that the senior partner would provide the associate with at least a day to construct a memo and probably a week to turn a memo into a brief that might be submitted to a court.²²⁵

Other people justify the speeded nature of the LSAT by saying it shows whether test takers have a different kind of speed—the ability to complete a project on a timely basis.²²⁶ One might argue that lawyers need to have the ability to meet important deadlines and that law students need to be able to study effectively to answer questions on an end-of-semester final exam. The proponents of this kind of speeded performance presume that the LSAT helps measure that kind of speeded skill. But that argument confuses speed with organizational skills. Meeting a court deadline or preparing effectively for a law school exam requires the organizational skill of working backward from a deadline and getting chunks of material completed on a periodic basis to meet the ultimate deadline. These are important planning skills. There is no reason to think that a slow reader has weak planning skills. One could imagine that slow readers have to compensate for the amount of time it takes them to read written work by having excellent planning skills, so they can use their time as efficiently as possible. The literature on speed emphasizes the complicated nature of that concept,²²⁷ yet lay responses to the need for speeded exams rarely reflect consideration of the multidimensional nature of speed.

(“Clients charged by billable hours also appreciate efficiency.”)

²²⁵ See Henderson, *supra* note 12, at 1035–36 (“An analysis hurriedly prepared in three hours may or may not be indicative of the work product that might arrive at the end of the day or the end of the week.”).

²²⁶ Practicing lawyers have made this argument to the author because it is important for lawyers to meet deadlines.

²²⁷ See *supra* Part II.A.

Even if one accepts the premise that it is valuable for lawyers to work relatively quickly, one needs to ask whether the speed required to do well on the LSAT is that kind of speed. Most test takers find the “analytical reasoning” section of the test, which is often called the “logic games” section,²²⁸ to be the most speeded section. In order to do well on that section, a test taker needs to learn certain shortcuts. The test developers do not provide test takers enough time to write out each permutation of the sets of rules that are provided. Although a test taker can probably learn those shortcuts to take the LSAT, it is not likely that an individual uses those particular shortcuts while in law school or as a practicing lawyer.²²⁹ Thus, it could be said that the test is measuring whether a test taker took the time to learn the shortcut, rather than measuring whether a test taker will be able to work quickly as a law student or lawyer.

Further, the fastest test taker may not necessarily be the best lawyer. In order to proceed quickly through a test such as the LSAT, a test taker needs to be willing to mark the right answer as soon as he or she spots it, proceeding through the A to E options. If A seems correct, the test taker needs to mark “A” and not even read the other options, but a good lawyer would rarely act so hastily. A good lawyer would want to read each option and double-check her answer in order not to make an error on behalf of a client. A good lawyer reads dense material word by word, looking for internal contradictions and inconsistencies. A good lawyer would wonder if more than one answer might actually be correct, rather than circling the first seemingly correct answer that is found. The “rush to complete” kind of speed embedded in all sections of the LSAT may actually be antithetical to the practice of law.²³⁰ But the LSAT, of course, only purports to predict performance in the traditional version of the first year of law school—a set of speeded exams—not the actual practice of law. The use of speeded exams in law school

²²⁸ One of the most popular test preparation services for the LSAT is PowerScore. Their books and website refer to the “analytical reasoning” section as the “logic games” section. See *LSAT Preparation Packages*, POWERSCORE, <https://www.powerscore.com/lsat/publications/> (last visited Nov. 3, 2018).

²²⁹ The author has been practicing LSAT questions through the Khan Academy site. As a highly successful lawyer and law professor, the author cannot successfully complete the logic games in the allocated time. She usually needs twice as much time as is provided to attain the correct answers. Thus, the author does not appear to have the skills needed to answer those questions quickly, although she completes legal tasks quite quickly. The speed tested in that section is not a speed that the author has attained even though she works quickly.

²³⁰ In the author’s decades of hiring law students to complete legal work for her, she has often had to counsel students to slow down and do careful work. Her fastest workers were rarely her best workers. And, of course, if LSAC were measuring the kind of speed that is useful in law practice, then those of us who have been working for decades as lawyers would find the LSAT to be an easy exam. That is a very improbable conclusion.

is not grounded in any formal or well defined theory. Rather, it is driven by a variety of nontheoretical factors, including the need to generate a grading curve, the desire to limit the volume of words that must be read, to enforce a consistent time limit, and the simple adherence to tradition.²³¹

The problem with highly speeded exams, such as the LSAT, is quite simple. Some people do not finish. For the questions they do not even reach, we have measured only their speed and not their skills, knowledge, and aptitude that the test purports to measure. We perpetuate the merit-speed myth by not insisting that test developers justify the validity of speeded conditions for the entire test-taking population. The ADA, by incorporating Title VII test-validity principles, provides us with an opportunity to undermine the merit-speed myth. It allows us to change who gets an affordable opportunity to sit in a law school classroom by requiring LSAC to abandon the speeded nature of its exam if it cannot validate the need for speed and demonstrate that a non-speeded exam would be a less-valid testing instrument. Unverified assumptions should not play such a prominent role in determining who sits in the law school classroom at an affordable price.²³²

VI. CONCLUSION

In this Article, the author has argued that the ADA should be used as a legal tool to require testing entities to allow all candidates to take tests under non-speeded conditions. Further, based on legal and equitable principles, educational programs could *instruct* testing entities, which are their clients, to develop exams that are less speeded. This suggestion relies on both Title VII test-validity and ADA universal-design principles. It is novel.

As a result of Pearson's leadership and the availability of college admissions without any standardized test scores, students can graduate from college without having faced speeded exams. We need to closely examine whether this change in our default expectations can continue up the ladder so that people like Elizabeth Black or Bernadette Bibber do not need to prove that they have a disability in order to receive enough time to finish a professional certification exam. Change can occur without any judicial intervention. Testing entities can abandon their assumption that faster is always better—it is not.

Beyond the world of standardized testing, this Article should also make all legal educators pause and ask about their own testing practices. How can they make changes consistent with universal design principles to avoid some

²³¹ Henderson, *supra* note 12, at 980.

²³² For a broad critique of the overuse of the LSAT as an admissions tool, see Phoebe A. Haddon & Deborah W. Post, *Misuse and Abuse of the LSAT: Making the Case for Alternative Evaluative Efforts and a Redefinition of Merit*, 80 ST. JOHN'S L. REV. 41 (2006).

students having to seek accommodations such as extended time?

When the author has given faculty workshops on this subject, she is often met with a roomful of faculty who ask that question. One faculty member, for example, asked why law schools presume that, when they set forth the exam schedule, that exams will be in a four-hour block. What if the default rule became eight hours (with strict word limits for essay exams) and faculty would have to ask for an exception to require an exam to be completed in less time? Changing our default rules about the value of speed could create dramatic changes in the educational system.

The current validity of the LSAT is based on its ability to predict grades in the first year of law school, where speeded exams have long been the norm. If professors who teach in the first year of law school stopped giving speeded exams, it is likely that that correlation would diminish. Thus, if law faculty change their own testing practices, then they, in turn, might put pressure on LSAC to change the structure of its exam.

LSAC loves to test candidates on their logical reasoning skills and make them spot unwarranted assumptions. So, consider the following future question for an LSAT exam:

LSAC has developed a test, called the LSAT, which is designed to predict first-year grades in law school. That test has been in existence for twenty-five years and has been repeatedly shown to meet professional standards of test validity. It is a four-hour, timed testing instrument. During that twenty-five year period, ninety-five percent of students enrolled in law school attained grades during their first year by taking four-hour, timed testing instruments, while five percent of students took essay exams with word limits rather than time limits. What assumption is being made by the professional testing entity:

- (a) That the current method of grading first-year law students will not change;*
- (b) That the validity of the LSAT is the same for the students who took timed and untimed first-year exams;*
- (c) That the practice of giving untimed law school exams in the first year of law school has not emerged since the last validity study of the LSAT was completed;*
- (d) All of the above;*
- (e) None of the above.*

If LSAC provides the correct answer—(d)—then it should reconsider the use of time limits on its exam. It would be best, however, if it answered that question quickly rather than wait another twenty-five years to figure out the right answer. In this context, faster would be better.