

January 2014

Testing, Diversity, *and* Merit: A Reply to Dan Subotnik and Others

Andrea A. Curcio

Carol L. Chomsky

Eileen Kaufman

Follow this and additional works at: <http://scholarship.law.umassd.edu/umlr>

 Part of the [Legal Education Commons](#), and the [Legal Profession Commons](#)

Recommended Citation

Curcio, Andrea A.; Chomsky, Carol L.; and Kaufman, Eileen (2014) "Testing, Diversity, *and* Merit: A Reply to Dan Subotnik and Others," *University of Massachusetts Law Review*: Vol. 9: Iss. 2, Article 1.

Available at: <http://scholarship.law.umassd.edu/umlr/vol9/iss2/1>

This Article is brought to you for free and open access by Scholarship Repository @ University of Massachusetts School of Law. It has been accepted for inclusion in University of Massachusetts Law Review by an authorized administrator of Scholarship Repository @ University of Massachusetts School of Law.

Testing, Diversity, *and* Merit: A Reply to Dan Subotnik and Others

Andrea A. Curcio, Carol L. Chomsky, and Eileen Kaufman

9 U. MASS. L. REV. 206

ABSTRACT

The false dichotomy between achieving diversity and rewarding merit frequently surfaces in discussions about decisions on university and law school admissions, scholarships, law licenses, jobs, and promotions. “Merit” judgments are often based on the results of standardized tests meant to predict who has the best chance to succeed if given the opportunity to do so. This Article criticizes over-reliance on standardized tests and responds to suggestions that challenging the use of such tests reflects a race-comes-first approach that chooses diversity over merit. Discussing the firefighter exam that led to the Supreme Court decision in *Ricci v. DiStefano*, as well as the LSAT and Bar Exam, the Article questions the way standardized tests are used in making critical gateway decisions. It argues, consistent with Title VII, that racially disparate test outcomes should prompt inquiry into whether better ways exist to determine merit. Based on studies indicating that cognitive tests predict academic and workplace success for a relatively small percentage of test-takers, and on research into assessing a wider range of skills in many fields, the Article suggests we can both better predict who will succeed as future lawyers *and* reduce the impact of test score racial disparities by modifying law school admissions and bar licensing processes. The Article concludes that questioning over-reliance on cognitive tests to measure merit will lead to the development of better assessment measures with more diverse outcomes, more fairness for all applicants, and more comprehensive decision-making processes that better reflect true merit.

AUTHORS

Andrea A. Curcio, Professor of Law, Georgia State University; Carol L. Chomsky, Professor of Law, University of Minnesota Law School; Eileen Kaufman, Professor of Law, Touro Law School. The authors have served as officers and board members of the Society of American Law Teachers (SALT) and have participated on behalf of SALT in activities related to the issues discussed in this Article and in developing statements made by SALT on these topics. This Article represents their individual views, not the views of the organization. The authors thank Ngai Pindell for his review of an earlier draft, Paul Petersen for his help with the research and citations, and the Georgia State University College of Law librarians for their help with the social science research.

2014	Testing, Diversity, and Merit	207
I.	INTRODUCTION.....	208
II.	RICCI V. DISTEFANO AND DISPARATE IMPACT	212
A.	The Appropriate Response to Disparate Impact	213
B.	What was Wrong with the <i>Ricci</i> Test?.....	217
III.	STANDARDIZED TESTS AND THE LEGAL PROFESSION.....	221
A.	Questioning the Bar Exam	222
B.	What’s Wrong with the Bar Exam?	230
1.	Bar Exam Format.....	231
2.	Focus on Memorization	232
3.	Impact of Test-Taking Speed.....	235
4.	Bar Exam Results and Law School Grading	239
5.	Bar Exam Results and Lawyering Competence.....	239
6.	What the Bar Exam Does Not Test.....	241
C.	The Bar Exam Constrains Change in Legal Education.....	242
IV.	ALTERNATIVES TO THE EXISTING BAR EXAM	244
A.	The New Hampshire Model.....	245
B.	Modify the Existing Exam	248
C.	Test More Skills	249
V.	CRITIQUING THE LSAT	252
A.	What’s Wrong with the LSAT?	253
B.	Learning from SAT-Optional Schools.....	255
C.	Broadening the Admissions Process	260
VI.	TEST-TAKING AND ECONOMICS	263
A.	Cognitive Tests and Job Performance.....	263
B.	The Role of Unconscious Bias in Workplace Performance Evaluations.....	266
C.	The Economic Value of Diversity	268
D.	OECD study.....	269
VII.	TEST-TAKING AND RACIAL GAPS.....	271
VIII.	CONCLUSION	275

I. INTRODUCTION

Beginning as early as 2002, commentators have urged re-examination of the use of standardized tests and the role they play in hiring, promotion, licensing, and admissions decisions because of the limited ability of those tests to predict academic and job success, combined with their disparate impact on racial minorities.¹ In *Does Testing = Race Discrimination?: Ricci, the Bar Exam, the LSAT, and the Challenge to Learning*,² Professor Dan Subotnik argues that such challenges to the reliance on standardized tests equate to a race-comes-first approach that chooses diversity over intellectual ability and economic growth.³ He suggests that the failure to use test scores as a

¹ See, e.g., Robert J. Sternberg, *The Rainbow Project: Enhancing the SAT through Assessments of Analytical, Practical, and Creative Skills*, 34 INTELLIGENCE 321 (2006) (arguing for augmenting the SAT with assessments that measure a wider range of skills) [hereinafter Sternberg, *Rainbow Project*]; Robert J. Sternberg and Jennifer Hedlund, *Practical Intelligence, g, and Work Psychology*, 15 HUMAN PERFORMANCE 143 (2002) (arguing that researchers and employers should look beyond general intelligence tests to predict successful job performance); Marjorie M. Shultz & Sheldon Zedeck, *Predicting Lawyer Effectiveness: Broadening the Basis for Law School Admission Decisions*, 36 LAW & SOC. INQUIRY 620 (2011) (suggesting ways to supplement the LSAT to account for a broader range of lawyering effectiveness factors); SALT, *Statement on the Bar Exam*, 52 J. LEGAL EDUC. 446 (2002) (arguing for re-examination of the bar exam); Kristin Booth Glen, *When and Where We Enter: Rethinking Admission to the Legal Profession*, 102 COLUM. L. REV. 1696 (2002) (arguing for re-thinking the bar exam); 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. JOHNS L. REV. 41 (2006) (arguing for re-evaluation of the LSAT as one of the main law school admissions criteria).

² Dan Subotnik, *Does Testing = Race Discrimination?: Ricci, The Bar Exam, the LSAT, and the Challenge to Learning*, 8 U. MASS. L. REV. 332 (2013).

³ See, e.g., *id.* at 398 (suggesting that those who critique standardized tests undermine the pursuit of knowledge and devalue intellectual achievement); *id.* at 344 (arguing that it hurts the economy when jobs go to people not best suited, and implying that test scores are a measure of who is best suited for a given job); *id.* at 346 (arguing that a “society that belittles knowledge and learning can pay a high economic price” and again implying that critiquing standardized test scores equates to belittling knowledge and learning); *id.* at 394 (arguing that concern over affirmative action and disparate impact results in “costs that are incurred not only by those who fail to get jobs for which they are better qualified but also by the entire society.”); *id.* at 395 (arguing that those who question the validity of the LSAT and Bar Exam as a gateway to the profession “pooh pooh the importance of learning and educational achievement in economic life”).

key determinant of who gets hired, promoted, admitted to universities and law schools, and licensed is unfair to individuals,⁴ economically damaging to society,⁵ and serves a postmodern vision that considers merit-testing as merely perpetuating the status of the powerful.⁶ Professor Subotnik's arguments repeat and reinforce commonly held beliefs about fairness and merit,⁷ but those arguments and beliefs are grounded in flawed assumptions and present a false dichotomy between the twin goals of achieving diversity and identifying qualified individuals. We write this response to address those flaws, to explicate a more nuanced view of the value of cognitive tests, and to support a broader, and more just, understanding of merit.

Standardized testing was introduced in the United States in an effort to ensure that hiring, promotion, admissions, and other important decisions about access to programs and jobs would be made based on objective criteria, not based on subjective judgments that could—and had been—infected by bias and discrimination and other irrelevant factors.⁸ Those who performed well on the tests were presumed to be

⁴ *Id.* at 339 (arguing that New Haven's refusal to certify job promotion tests that had a disparate impact was unfair to the white firefighters who took the test); *id.* at 381 (arguing that moving from an admissions standard that looks at individual LSAT scores to one that looks at score ranges is "unfair at the individual level").

⁵ *See id.* at 394 (arguing that failure to rely upon cognitive tests to make hiring decisions and failure to focus on increasing cognitive test scores is economically harmful).

⁶ *Id.* at 354–57.

⁷ *See* Lani Guinier, *The Future of Affirmative Action, Reclaiming the American Ideal*, 84 CAL L. REV. 953, 960–61 (1996). Guinier notes that:

[the] stock story of affirmative action critics in the employment context (and the one that appears most often in the cases) is of the white civil servant—say a police officer or firefighter—John Doe. He scored several points higher on the civil service exam and interview rating process, but lost out to a woman or person of color who did not score as high on those selection criteria. John Doe claims, along with many public opponents of affirmative action, that he is more qualified for the job, and that it is unfair to allow race or gender considerations to deprive him of what he 'deserves.'

Id.

⁸ Lani Guinier, *Admissions Rituals as Political Acts: Guardians at the Gates of our Democratic Ideas*, 117 HARV. L. REV. 113, 131–32 (2003); Soohan Kim, Alexandra Kalev, & Frank Dobbin, *Progressive Corporations at Work: The Case of Diversity Programs*, 36 N.Y.U. REV. L. & SOC. CHANGE 171, 192 (2012); Haddon & Post, *supra* note 1, at 45. Although the general understanding

best qualified and therefore to deserve placement in gifted programs, entry into elite schools, and the award of scholarships, jobs and promotions.⁹ Despite such laudable goals, however, standardized test results are far from infallible predictors of who is likely to succeed.¹⁰ Too much reliance on test scores to measure merit discounts or ignores the fact that standardized tests measure only one aspect of intelligence,¹¹ are correlated with income and parental education,¹² and produce results that may be skewed because of contextual cues related more to outcome expectations than to actual ability.¹³

is that standardized testing was adopted to avoid bias, as noted in the text, at least one commentator has expressed a different view, suggesting that tests like the SAT were adopted to “create[] a path to upward mobility and national leadership for intelligent men of middle class means” while allowing continued exclusion of applicants of different gender and racial and ethnic backgrounds. Kimberly West-Faulcon, *More Intelligent Design: Testing Measures of Merit*, 13 U. PA. J. CONST. L. 1235, 1260–61 (2011).

⁹ Guinier, *supra* note 8, at 132; *see also* Michael Selmi, *Understanding Discrimination in a “Post-Racial” World*, 32 CARDOZO L. REV. 833, 851 (2011) (noting that in *Ricci*, the Court “viewed test results as consistent with their expectations: to the Court, it was to be expected that the white firefighters would perform better than the minority firefighters and the results were explained by their hard work and superiority rather than by problems with the test”).

¹⁰ *See infra* Parts V.A–B, VI.A (discussing the limited ability of standardized tests to predict academic success and job performance).

¹¹ *See infra* Parts V.C, VI.A (discussing various aspects of intelligence beyond cognitive intelligence and how those forms of intelligence also help predict academic and job success).

¹² *See College Admissions Show Test Driven Schooling Fails*, FAIRTEST EXAMINER (Fall, 2013) <http://www.fairtest.org/college-admissions-tests-show-testdriven-schooling> (compiling data on SAT test scores by family income); Joseph A. Soares, *Private Paradigm Constrains Public Response to Twenty First Century Challenges*, 48 WAKE FOREST L. REV. 427, 436–37 (2013) (discussing evidence of significant correlation between SAT scores and parental income while finding little correlation between high school GPA and parental income and arguing that admissions based upon test scores is a “form of Social Darwinism with social selection for high income families disguised as academic selection for the best talent”); Lucille A. Jewel, *Merit and Mobility*, 43 U. MEM. L. REV. 239, 270 (2012) (noting that the “children of college educated parents score 150 points higher on the SAT than children whose parents are high-school dropouts”).

¹³ Claude M. Steele & Joshua Aronson, *Stereotype Threat and the Intellectual Test Performance of African Americans*, 69 J. PERSONALITY & SOC. PSYCHOL. 797 (1995) (discussing findings that perception of racial stereotypes can affect standardized test performance of African American students). Steele’s work has

Because standardized tests are an imperfect measure of ability and success, we must learn to look beyond performance on traditional standardized tests as we conceptualize merit and allocate important human capital resources. Abilities and attributes beyond those measured by standardized tests should be part of the equation. Research indicates we can test more than cognitive ability and when we do so, we increase the ability to predict who will succeed. When we give undue weight to cognitive test scores, we may award jobs and admissions slots to those who may not actually be the best qualified or we may exclude those who are equally capable of academic and job performance but who do not perform well on standardized tests.¹⁴

And what about achieving diversity, which Professor Subotnik suggests is the result-oriented basis for challenges to standardized testing? Concern about racially skewed test outcomes is indeed a reason why some have questioned the legitimacy of standardized tests, but such questioning is entirely appropriate, especially if it results in better tests as well as more diverse—and more fair—outcomes. Achieving fair and appropriate decision-making—not achieving racial balance—is the goal, but racial imbalance in outcomes is reason enough to explore the validity of the tests.

In this Article, we address the misconception that standardized tests should be the key determinant in deciding who gets an admissions slot, a job or promotion, or admitted into law practice. We do not advocate ignoring such test results, but instead suggest that it is unfair and unreasonable to place undue weight upon tests that do not measure the full range of relevant abilities and are considerably less predictive than assumed. We posit that the limited predictive ability of standardized tests, combined with their disparate impact, is reason to re-examine the tests and search for better alternatives.

We begin, as Professor Subotnik did, with a discussion of *Ricci v. DiStefano*,¹⁵ a case in which white firefighters sued the City of New Haven for its failure to certify the results of a job promotion exam that had a disparate impact. *Ricci* serves as the foundation upon which

been both widely cited and validated. For an overview of the literature on stereotype threat, see Clark D. Cunningham et al., *Passing Strict Scrutiny: Using Social Science to Design Affirmative Action Programs*, 90 GEO. L. J. 835, 839 (2002).

¹⁴ See *infra* Parts V.B, VI.A (discussing studies which indicate standardized test scores alone are not the best predictors of academic or job performance success).

¹⁵ 557 U.S. 557 (2009).

Professor Subotnik builds his argument that our society values “diversity above qualifications.”¹⁶ In Part II, we offer a different vision of *Ricci*, supported by the law of Title VII and the facts on the ground in New Haven. We explain why *Ricci* is a case in which the employer, faced with disparate impact, made the reasonable decision to explore the validity of the challenged test and the availability of better methods of assessing who was most qualified to perform the job. In Parts III and IV, we challenge Professor Subotnik’s conclusions about the appropriate response to the disparate impact of bar examination results, exploring the validity of the current test and the availability of alternatives. We suggest that the limited predictive abilities of the bar exam, combined with its disparate impact, present cogent reasons to examine whether better assessment alternatives exist, and we discuss several alternatives that may be, or have been, adopted. In Part V, we consider similar questions with respect to the Law School Admission Test. In Part VI, we respond to Professor Subotnik’s claims that challenging and supplementing traditional cognitive tests undermines economic development. Finally, in Part VII, we address directly the racial performance gap—the disparate impact that leads to the very different conclusions that Professor Subotnik and we reach. We conclude as we began, with our assertion that questioning the validity and comprehensiveness of the bar exam, the LSAT, and cognitive job tests is not anti-intellectual, as he suggests.¹⁷ Rather, exploration of better assessment methods embodies the truly intellectual approach to the question of who, in a society with limited resources, should reap its rewards.

II. RICCI V. DISTEFANO AND DISPARATE IMPACT

Professor Subotnik uses the 2009 Supreme Court case of *Ricci v. DiStefano* as the foundation for his discussion of the “legal, political, and moral challenges to testing”¹⁸ and as a prime example of his claim that those who question the legitimacy of tests that have a disproportionate negative impact upon people of color believe that

¹⁶ Subotnik, *supra* note 2, at 339; *id.* at 344.

¹⁷ *Id.* at 369–70 (arguing that SALT’s critiques of the bar exam equate to favoring diversity over learning); *id.* at 395 (arguing that those who question the validity of the LSAT and Bar Exam as a gateway to the profession “poooh poooh the importance of learning and educational achievement in economic life”).

¹⁸ *Id.* at 347.

“racial balance is the axis upon which our system of justice must turn.”¹⁹ Noting that the City would not have thrown out the test if there had not been a disparate impact on firefighters of color,²⁰ Professor Subotnik argues that New Haven invalidated its own tests in order to produce “black winners”²¹ and that doing so unfairly penalized mostly white firefighters who studied hard to pass the test. Thus, according to Professor Subotnik, *Ricci* illustrates the principle that some people value “diversity above all.”²²

We, too, begin with a discussion of *Ricci*, because the understanding—and misunderstanding—of that case is fundamental to the discussion of what it means to challenge the use of tests that produce race-based gaps in outcomes. *Ricci* is not, as Professor Subotnik claims, a case of “diversity above all.” Rather, it is a case that demonstrates a very real struggle to ensure that a job promotion test that produces a disparate impact is, in fact, both valid and the best available measure of job qualifications. Whether one agrees with the outcome in *Ricci*—and four Justices did not, a nuance lost in Professor Subotnik’s description—the circumstances that led to the litigation should be seen as an example of grappling with the hard issues surrounding the use of standardized testing, not as an illustration of choosing diversity over merit.

A. The Appropriate Response to Disparate Impact

The *Ricci* decision reviewed the actions of the City of New Haven after it administered promotion examinations for fire department supervisor positions. African American firefighters made up thirty percent of the City’s firefighters and nine percent of those ranked captain and above.²³ The promotion test results produced significant and unexpected racial disparities²⁴ and, if certified, would have resulted in no African American members of the department being eligible for promotion to either lieutenant or captain.²⁵ Because such a disparate impact can result in Title VII liability if the employer cannot

¹⁹ *Id.* at 339.

²⁰ *Id.* at 332.

²¹ *Id.* at 339.

²² *Id.* at 353.

²³ *Ricci v. DiStefano*, 557 U.S. 557, 610–11 (2009) (Ginsburg, J., dissenting).

²⁴ Brief for Respondent at 4, *Ricci v. DiStefano*, 557 U.S. 557 (2009).

²⁵ *Id.*

show that the test is job-related and a matter of business necessity and rebut any showing by the plaintiff that there are less racially disparate alternatives,²⁶ the City embarked upon an investigation to determine whether the test was a valid measure of firefighting leadership and whether there were equally valid, or better, ways to test for the positions that would have less of an adverse impact.²⁷

The City's investigation produced evidence that such better alternatives did exist.²⁸ Based upon the results of its investigation, the City refused to certify the test results.²⁹ Frank Ricci, seventeen other white firefighters, and one Hispanic firefighter, all of whom had achieved high scores on the promotion tests, filed suit alleging that negating the test results was a racially discriminatory act against the white firefighters in violation of the disparate treatment provisions in Title VII of the Civil Rights Act, which forbids intentional discrimination.³⁰ The lawsuit thus claimed that the City's actions undertaken to avoid liability for disparate *impact* resulted in disparate *treatment* forbidden by Title VII.

In a 5-4 ruling, the Court decided in favor of the plaintiffs, finding that the refusal to certify the test results violated Title VII's disparate treatment prohibition despite the tests' disparate impact.³¹ In the view

²⁶ The city was concerned about disparate impact liability, which arises not based on acts of intentional discrimination (though those also are forbidden) but based on unexplained or unjustified disparities in impact based on race, color, religion, sex, or national origin. Plaintiffs may establish a prima facie case of disparate impact by demonstrating that a facially neutral standard has a disproportionately adverse effect on minorities. Once plaintiffs establish a prima facie case, the burden shifts to the defendant to "demonstrate that the challenged practice is job related for the position in question and consistent with business necessity." 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012). If the defendant establishes the practice or test is job related and consistent with business necessity, plaintiffs may still prevail on a disparate impact claim if they can prove that a viable alternative existed that had less of a discriminatory impact and that the employer failed to adopt that alternative. 42 U.S.C. § 2000e-2(k)(1)(A)(ii)(2012). Disparate impact analysis was first established in *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971) and has since been enacted into the text of Title VII itself.

²⁷ See *supra* discussion of the legal standard in note 26. For a discussion of the city's actions, see *infra* text accompanying notes 44–62.

²⁸ *Ricci*, 557 U.S. at 572.

²⁹ *Id.* at 572–74.

³⁰ *Id.* at 575.

³¹ See *id.* at 583. The majority and dissent in *Ricci* disagreed about the appropriate legal rule to use in resolving the conflict created by the statute's mandates to

of the majority, the tests were clearly job-related and consistent with business necessity under the analysis applicable to disparate impact claims, and the City had an insufficient evidentiary basis to establish that there were equally valid, less discriminatory alternatives that served the City's needs.³² Justice Ginsburg, writing for the four dissenters, disagreed with these evidentiary findings, believing that the majority "ignores substantial evidence of multiple flaws in the tests New Haven used. The Court similarly fails to acknowledge the better tests used in other cities, which have yielded less racially skewed outcomes."³³

In *Ricci*, the City questioned the test's validity as the decisive factor in promotion because of the disparate impact of the test results. Professor Subotnik objects to "the peculiar circumstance that [New Haven] invalidated its own test for reasons of race."³⁴ Yet Title VII is designed to produce exactly that result—to cause remedial action to be taken, either voluntarily or compelled by a court judgment, when a seemingly neutral practice or policy leads to a disparate racial impact that is not justified. Federal regulations make this clear: "Whenever the user is shown an alternative selection procedure with evidence of less adverse impact and substantial evidence of validity for the same job in similar circumstances, the user should investigate it to determine the appropriateness of using or validating it in accord with these guidelines."³⁵ Examination of conduct that may violate a legal

avoid both kinds of discrimination. The majority thought that an employer should only be able to avoid disparate treatment liability if it could show that it had a "strong basis in evidence" that, without such action, it would be liable because of the disparate impact of the test results. *Id.* The dissenters thought an employer should only have to show that it had good cause to believe the test that produced the disparate impact would not be allowable as a business necessity. *Id.* at 625 ("an employer who jettisons a selection device when its disproportionate racial impact becomes apparent does not violate Title VII's disparate-treatment bar automatically or at all, subject to this key condition: The employer must have good cause to believe the device would not withstand examination for business necessity.").

³² *Id.* at 585, 587.

³³ *Id.* at 608–09 (Ginsburg, J., dissenting).

³⁴ Subotnik, *supra* note 2, at 347.

³⁵ 29 C.F.R. § 1607.3. This regulation states:

[W]henever a validity study is called for by these guidelines, the user should include, as a part of the validity study, an investigation of suitable alternative selection procedures and suitable alternative

prohibition against discrimination is exactly what disparate impact legal rules mandate.³⁶ Concern about disparate impact liability has played a significant role in raising employers' awareness of the need to examine, and potentially reform, employment practices that have "inhibited the economic advance of women and minorities."³⁷ Investigating the validity of a test because it has an unintentional disparate impact does not equate to having racial balance be the "axis on which our system of justice must turn."³⁸ Rather, it is recognition that when an examination produces an unintentional racial disparity,³⁹ it should prompt scrutiny of that test and a review of viable alternatives. As the New Haven City attorney noted, "significant adverse impact . . . triggers a much closer review [of the test], because it's like setting off a warning bell that there may be something wrong."⁴⁰

methods of using the selection procedure which have as little adverse impact as possible, to determine the appropriateness of using or validating them in accord with these guidelines. If a user has made a reasonable effort to become aware of such alternative procedures and validity has been demonstrated in accord with these guidelines, the use of the test or other selection procedure may continue until such time as it should reasonably be reviewed for currency. Whenever the user is shown an alternative selection procedure with evidence of less adverse impact and substantial evidence of validity for the same job in similar circumstances, the user should investigate it to determine the appropriateness of using or validating it in accord with these guidelines. This subsection is not intended to preclude the combination of procedures into a significantly more valid procedure, if the use of such a combination has been shown to be in compliance with the guidelines.

Id.

³⁶ *Id.*

³⁷ Lawrence Rosenthal, *Saving Disparate Impact*, 34 CARDOZO L. REV. 2157, 2158 (2013).

³⁸ Subotnik, *supra* note 2, at 339.

³⁹ On the captain exam, the white candidate pass rate was 64% compared to a 37.5% pass rate for both black and Hispanic test-takers. On the lieutenant exam, white candidates had a pass rate of 58.1%; black candidates had a pass rate of 31.6%; and Hispanic candidates had a pass rate of 20%. *Ricci v. DiStefano*, 557 U.S. 557, 586 (2009).

⁴⁰ Brief for Respondent, *supra* note 24, at 7.

When the test's results indicated a disparate impact, the City should have acted exactly as it did: scrutinized the test to ensure that the test was the best available measure of potential success in the firefighting leadership positions. This scrutiny not only comports with the legal standard,⁴¹ it comports with the need to ensure the testing process itself was fair and resulted in the best possible firefighter leaders.⁴² The fact that some outside groups pressed the City to respond to the test's disparate impact, highlighted by the majority opinion and by Professor Subotnik,⁴³ does not mean the City acted inappropriately in attending to those concerns. While the Supreme Court decided the City had insufficient basis for acting—a conclusion rejected by four Supreme Court Justices—the City cannot be charged with discarding the test simply because the outcome was not what it hoped it would be, as Professor Subotnik claims.

B. What was Wrong with the *Ricci* Test?

What did the City of New Haven do, and what should any employer do, when confronted with racial gaps in test results? Both to ensure fairness and to forestall legal liability, an employer offering rewards—hiring, promotion, or other benefits—on the basis of test outcomes must ensure that the tests are valid and reliable measures, that they measure the qualifications necessary, and that there are no alternative measures that do both of those tasks without creating significant racial imbalances.

That is precisely the inquiry made by the City of New Haven.⁴⁴ In *Ricci*, there was evidence presented to the City and to the trial court

⁴¹ Title VII prohibits employment practices that have a discriminatory impact if better assessment alternatives are available. *See supra* note 26.

⁴² As the *Ricci* dissent notes, “[f]irefighting is a profession in which the legacy of racial discrimination casts an especially long shadow. . . . The [U.S. Commission on Civil Rights] Report singled out police and fire departments for having ‘[b]arriers to equal employment . . . greater . . . than in any other State or local government’” *Ricci*, 557 U.S. at 609–10 (Ginsburg, J., dissenting). Diversity in the firefighting force helps develop firehouse camaraderie, promotes sharing of information, tolerance, and mutual respect among colleagues, and builds knowledge of diverse communities within the city. *See Lomack v. City of Newark*, 463 F.3d 303, 309 (2006). The City's tests failed to assess many of the skills and qualities that indicate who will be the best firefighting lieutenants and captains. *See infra* text accompanying notes 48–55.

⁴³ *See Ricci*, 557 U.S. at 598; Subotnik, *supra* note 2, at 339.

⁴⁴ *Ricci*, 557 U.S. at 613 (Ginsburg, J. dissenting).

that the exams were *not* the best measure of who would be the most qualified firefighter lieutenants and captains, and that more accurate, and less discriminatory, assessment alternatives *did* exist.⁴⁵ For example, one key deficiency in the promotion exams was their failure to assess command presence and the ability to quickly assess and respond to changing and often confusing events.⁴⁶ During a fire, leaders must be able “to act decisively, to communicate orders clearly and thoroughly to personnel on the scene, and to maintain a sense of confidence and calm even in the midst of intense anxiety, confusion and pain.”⁴⁷ The New Haven exams did not attempt to measure command presence; they were designed to measure only job-related knowledge.⁴⁸ This failure to measure a critical skill meant that “a high test score thus could not support an inference that the candidate would be a good commander in the line of duty; conversely, those candidates with strong command attributes were never given an opportunity to demonstrate them.”⁴⁹ The New Haven tests also failed to measure proficiency in interpersonal relations, supervisory skills, and the ability to function under dangerous circumstances.⁵⁰ While it is true that knowledge of fire science is important to a commander, the highest scorer on the exam is not necessarily the best qualified firefighting leader. As the City attorney noted, “the goal of the test is to decide who is going to be a good supervisor ultimately, not who is going to be a good test-taker.”⁵¹

There was evidence that some of the deficiency in the scope of the multiple choice portion of the test could have been addressed by

⁴⁵ See Brief for Industrial Organizational Psychologists as Amicus Curiae Supporting Respondents at 12, *Ricci v. DiStefano*, 557 U.S. 557 (2009) (Nos. 07-1428, 08-328), 2009 WL 795281 (“Virtually all studies of fire management emphasize that command presence is vital to the safety of firefighters at the scene and to the successful accomplishment of the firefighting mission and the safety of the public.”) [hereinafter IOP Brief].

⁴⁶ See *id.*; Richard B. Gasaway, *Making Intuitive Decisions Under Stress: Understanding Fireground Incident Command Decision-Making*, 1 INT’L FIRE SERV. J. OF LEADERSHIP & MGMT. 8 (2007).

⁴⁷ IOP brief *supra* note 45, at 11 (citing RICHARD KOLOMAY & ROBERT HOFF, *FIREFIGHTER RESCUE & SURVIVAL* 5–13 (2003)).

⁴⁸ *Id.*

⁴⁹ *Id.* at 12.

⁵⁰ See *Ricci v. DiStefano*, 557 U.S. 557, 634 (2009) (Ginsburg, J., dissenting).

⁵¹ IOP brief, *supra* note 45, at 12.

weighing the oral exam more heavily—60% oral and 40% written rather than the reverse, as it was weighed.⁵² The over-emphasis on the multiple choice questions was exacerbated by the admittedly arbitrary seventy percent cut-off passing score required under the City’s civil service rules⁵³ and by the decision of the City’s test consultant, Industrial/Organizational Solutions, Inc. (IOS), to design more difficult test questions to screen out more people.⁵⁴ These choices skewed the test toward the “attenuated set of knowledge and abilities that are measured by a multiple choice test” and increased the adverse impact on minority candidates by ignoring other knowledge, skills, attitudes, and other personal characteristics needed to be a fire officer.⁵⁵

If the test—while reliable and valid to evaluate the narrow set of traits it set out to measure—did not accurately identify the best candidates for promotion, were alternative tests available that could do so, and without a significant racial gap in outcomes? The answer, based on experience in other communities, was yes.⁵⁶ Because of the problems with paper-and-pencil tests, over two-thirds of the country’s municipalities rely on other testing methods, including the use of assessment centers.⁵⁷ Rather than employing oral or multiple choice exams, assessment centers use simulations of real-world situations that require test-takers to demonstrate how they would address the problem in real life⁵⁸ and employ job simulations to test command presence.⁵⁹ For those municipalities that still partly relied on written exams, “the median weight assigned to them was 30 percent—half the weight

⁵² See Cheryl I. Harris & Kimberly West-Faulcon, *Reading Ricci: Whitening Discrimination, Racial Test Fairness*, 58 UCLA L. REV. 73, 151 at Table 5 (2010). The arbitrary weighting of test sections affected white, as well as minority, promotion candidates. *Id.* at 133–35.

⁵³ IOP brief, *supra* note 45, at 17.

⁵⁴ *Id.*

⁵⁵ *Id.* at 18–19.

⁵⁶ See *Ricci v. DiStefano*, 557 U.S. 557, 634–35 (2009) (Ginsburg, J., dissenting).

⁵⁷ *Id.*

⁵⁸ *Id.* at 570–71 (2009).

⁵⁹ IOP brief, *supra* note 45, at 31–32 (citing Diana E. Krause et al., *Incremental Validity of Assessment Center Ratings Over Cognitive Ability Tests: A Study at the Executive Management Level*, 14 INT’L J. SELECTION & ASSESSMENT 360, 362 (2006)).

given to New Haven's written exam."⁶⁰ The New Haven Civil Service Board heard testimony about these alternative methods and learned that these methods "were both more reliable and notably less discriminatory" than the process used by New Haven.⁶¹ As the dissent in *Ricci* noted, "A test fashioned from materials pertaining to the job . . . superficially may seem job-related. But the issue is whether it demonstrably selects people who will perform better [all of] the required on-the-job behaviors."⁶²

Despite evidence that the New Haven tests were seriously flawed as a measure of fire officer qualifications—enough to convince the four dissenters—the *Ricci* majority concluded that discarding the test results was unfair to those with high scores because it upset their "legitimate expectations" of promotion if they performed well on the test.⁶³ Professor Subotnik reaches a similar conclusion.⁶⁴ This reasoning is backwards. Legitimate expectations depend upon a legitimate selection method.⁶⁵ As Justice Ginsburg stated, "If an employer reasonably concludes that an exam fails to identify the most qualified individuals and needlessly shuts out a segment of the applicant pool, Title VII surely does not compel the employer to hire or promote based on the test, however unreliable it may be."⁶⁶

In hindsight, no one disputes that it would have been preferable for the City to have more expansively investigated testing methods prior to administering the test or to have given the test company authority to explore a full range of testing rather than constraining them to work within narrower parameters.⁶⁷ It was unfair to the firefighters who studied for a test that was not the best measure of qualification for promotion to later dash their expectations when they scored highly on that test. But it was also unfair to firefighters who did not score highly

⁶⁰ *Ricci*, 557 U.S. at 635 (Ginsburg, J., dissenting).

⁶¹ *Id.*

⁶² *Id.* at 637 (quoting *Boston Chapter NAACP v. Beecher*, 504 F.2d 1017, 1021–22 (1st Cir. 1974)).

⁶³ *Id.* at 583.

⁶⁴ See Subotnik, *supra* note 2, at 401.

⁶⁵ See *Ricci v. DiStefano*, 557 U.S. 557, 630 (2009) (Ginsburg, J., dissenting).

⁶⁶ *Id.*

⁶⁷ *Id.* at 637 (Ginsburg, J., dissenting).

on the existing test⁶⁸ to administer and use for promotions a test that was significantly flawed as a measure of job qualification, especially in light of the availability of better and less discriminatory tests.

Professor Subotnik argues not just that the City of New Haven made a judgment based on race but that, in doing so, it put the needs of the minority over the needs of the majority and failed to account for the costs of “ignoring job preparedness.”⁶⁹ While it is certainly possible for a decision-maker to do that, Professor Subotnik paints with too broad a brush. Motivated by the desire to avoid unjustified disparate results, the decision-makers looked at what the City truly needed in qualified leaders for the fire department. They decided to ignore the results of what they determined was an inferior and incomplete process. The City of New Haven had an interest in having the most qualified fire officers leading its department and an interest in having tests that best identified those people. As Professors Harris and West-Faulcon note, “to the extent disparate impact law pushes employers to make actual merit-based employment decisions, all racial groups, individual applicants, and society as a whole benefit.”⁷⁰

III. STANDARDIZED TESTS AND THE LEGAL PROFESSION

Ricci involved the appropriate response when standardized testing of firefighters seeking promotions resulted in significant differences in outcome with respect to race, but the story told and the issues raised and decided in that case have broad implications for all use of standardized testing that produces such gaps. While *Ricci* is the starting place, Professor Subotnik’s primary targets are those who challenge the use of standardized testing as the key determinant for entry into law school and the legal profession. Professor Subotnik suggests that questioning reliance on the LSAT and bar exam because of the tests’ disparate impact equates to denigrating knowledge and intelligence: “a rich irony is served up when legal academics pooh-pooh the importance of learning and educational achievement in economic life.”⁷¹ He further suggests that if a decision-maker really

⁶⁸ Harris & West-Faulcon, *supra* note 52, at 127 (noting that white firefighters were also harmed by the exam).

⁶⁹ Subotnik, *supra* note 2, at 392–94.

⁷⁰ Harris & West-Faulcon, *supra* note 52, at 121.

⁷¹ Subotnik, *supra* note 2, at 395. Throughout his article, Professor Subotnik makes this “knowledge or diversity” point in numerous ways: e.g. jobs go to

wants to ensure inclusion of minorities, it can administer “simplified” questions or set a low threshold for qualification and then use a lottery for selection.⁷² That insulting argument⁷³ rests on the false premise that the tests validly measure what it takes to succeed, so achieving diversity requires “moving the finish line”⁷⁴ to designate the winners irrespective of merit. Professor Subotnik thereby suggests a false dichotomy in law as in fire-fighting: best qualified applicant or diverse outcome, but not both. As argued in the previous section, tests that produce disparate outcomes should be examined to ensure that they are, in fact, valid measures of job or academic performance and that there are no viable, less discriminatory alternatives. Professor Subotnik clearly believes that the LSAT and the bar exam are appropriate gateways to law school, to law licensing, and to law practice. In this section, we challenge that assumption and describe how to better protect the public interest by testing more comprehensively the skill set necessary to succeed in law.

A. Questioning the Bar Exam

In 2002, the Society of American Law Teachers (SALT) published a critique of the bar exam, suggesting that the exam poorly measures who has minimum competence to practice law.⁷⁵ Just as unexpected disparate outcomes prompted the City of New Haven to examine its firefighting promotion tests to ensure they were the best predictors of

people who are not best suited. *Id.* at 344; OCED study suggests that a society that belittles knowledge and learning can pay a high economic price. *Id.* at 346; race again, above all else. *Id.* at 370.

⁷² Subotnik, *supra* note 2 at 347.

⁷³ It is not the suggestion of using a lottery for selection that is insulting. As Professor Subotnik notes, Susan Sturm and Lani Guinier have also suggested using a lottery for selection. However, their point is quite different and is based on the belief that the tests are *not* valid as selection tools and that the opportunities offered (of a job, a promotion, a place in an educational institution) should be distributed more randomly—and therefore equitably—among “relatively indistinguishable candidates.” See Susan Sturm & Lani Guinier, *The Future of Affirmative Action: Reclaiming the Innovative Ideal*, 84 CAL. L. REV. 953, 1012 (1996). However, Professor Subotnik connects examination performance with quality so “lowering the threshold” means, to him, lowering the quality of those receiving the benefit or reward in the service of “race above all.”

⁷⁴ Subotnik, *supra* note 2, at 338.

⁷⁵ See *SALT Statement on the Bar Exam*, *supra* note 1.

job success, the bar exam's disparate outcomes⁷⁶ prompted SALT to encourage states to review their bar exams to determine whether those exams were the best predictive measure of future lawyers' competence and whether viable and less discriminatory alternatives existed. Just as the City of New Haven sought fair measures that could demonstrate who would make the best lieutenants and captains, SALT sought, and continues to seek, fair outcomes that can demonstrate who will make successful and competent lawyers.

Like those critiquing the New Haven firefighters' exam as having an insufficient relationship to how firefighting leaders employ their skills at the stationhouse and in actual fires, SALT critiqued the bar exam based upon the disjunction between what the bar exam tests and how lawyers practice.⁷⁷ Professor Subotnik and other defenders of the bar exam argue that even though the bar exam fails to test all the skills lawyers need, the skills it does test—reading comprehension, issue spotting, legal reasoning and analysis, and written communication⁷⁸—

⁷⁶ LINDA F. WIGHTMAN, LSAC NATIONAL LONGITUDINAL BAR PASS STUDY 27 (1998), <http://www.unc.edu/edp/pdf/NLBPS.pdf> (reporting findings that first-time bar pass rates were 92% for whites, compared to 61% for African Americans, 66% for Native Americans, 75% for Mexican Americans/Hispanics, and 81% for Asian American).

⁷⁷ See *SALT Statement on the Bar Exam*, *supra* note 1.

⁷⁸ According to the National Conference of Bar Examiners [NCBE], the multiple choice and essay question portion of the exam tests the ability to:

- (1) identify legal issues raised by a hypothetical factual situation;
- (2) separate material which is relevant from that which is not; (3) present a reasoned analysis of the relevant issues in a clear, concise, and well-organized composition; and (4) demonstrate an understanding of the fundamental legal principles relevant to the probable solution of the issues raised by the factual situation. The primary distinction between the MEE [Multistate Essay Examination] and the Multistate Bar Examination (MBE) is that the MEE requires the examinee to demonstrate an ability to communicate effectively in writing.

Overview of the MEE, NAT'L CONF. OF BAR EXAMINERS, <http://www.ncbex.org/about-ncbe-exams/mee/overview-of-the-mee/> (last visited June 18, 2014).

The Multistate Performance Test is described as measuring the ability to:

- (1) sort detailed factual materials and separate relevant from irrelevant facts; (2) analyze statutory, case, and administrative materials for applicable principles of law; (3) apply the relevant

are so foundational that one cannot be an effective lawyer without minimum competence in those skills.⁷⁹ Of course, lawyers must be able to read, spot issues, and engage effectively in legal analysis and written communication. But saying those skills are fundamental does not establish that the bar exam is the best way to assess baseline competence in light of how those skills are used by lawyers, or that it is appropriate to use a test of only those skills as a gateway to the profession. Nor does it address whether there are viable, and better, ways to test for lawyering skills, and without a disparate racial impact.

Professor Subotnik focuses his attention on SALT's critique of the bar exam, but similar concerns have been raised for years, not only by other commentators, but by the organized bar, which has repeatedly questioned the traditional bar exam because of the narrow range of skills it tests and the disparate racial impact it produces. The New York organized bar's decades-long history of studying the bar exam is illustrative and worth describing at some length to underscore that

law to the relevant facts in a manner likely to resolve a client's problem; (4) identify and resolve ethical dilemmas, when present; (5) communicate effectively in writing; and (6) complete a lawyering task within time constraints.

Overview of the MPT, NAT'L CONF. OF BAR EXAMINERS, <http://www.ncbex.org/about-ncbe-exams/mpt/overview-of-the-mpt/> (last visited June 18, 2014).

⁷⁹ See Subotnik, *supra* note 2, at 371; Suzanne Darrow-Kleinhaus, *A Response to The Society of American Law Teachers Statement on the Bar Exam*, 54 J. LEGAL EDUC. 442, 442 (2004) (noting that the bar exam tests "reading comprehension and reasoning, identifying and formulating legal issues, organizing information, following directions and the ability to write"); Denise Riebe, *A Bar Review for Law School: Getting Students on Board to Pass Their Bar Exams*, 45 BRANDEIS L. J. 269, 279 (2006). Riebe argues that:

[B]ar exams test many fundamental skills which should have been learned in law school and which are essential to the practice of law, including: reading critically, comprehending what is read, reasoning logically, analyzing factual scenarios, separating relevant from irrelevant information, mastering and understanding legal rules, performing under time constraints, meeting time deadlines, identifying legal issues, applying legal rules to clients' situations, thinking like a lawyer, organizing information, following directions, and communicating effectively in writing.

Id.

challenges to the bar exam are not politically motivated, nor do they represent a “race above all” attitude.

As far back as 1992, the Committee on Legal Education of the New York City Bar Association identified problematic aspects of the bar exam and expressed the view that “the NYS Bar Exam does not adequately or effectively test minimal competence to practice law in New York.”⁸⁰ In its report, the committee recommended reducing the number of doctrinal areas tested, assessing more lawyering skills, eliminating all multiple choice questions (including the Multistate Bar Exam as well as multiple choice questions based on New York law), and eliminating the Multistate Professional Responsibility Exam.⁸¹ The Committee suggested substituting new performance-test questions for the multiple choice questions, integrating ethical issues into the essays or the performance test, and exploring the use of videotape examination questions for both skills and doctrinal questions.⁸² The report raised concerns about the disproportionate effect of the bar exam on minority applicants, which it found problematic on public policy grounds, especially in light of the report’s findings regarding the inadequacies of the exam.⁸³ It recommended that data be collected to assess disparate impact and that all revisions to the bar exam be evaluated in light of whether they would exacerbate that problem. The goal, the committee said, should be to reduce the disproportionate impact while enforcing reasonable standards of attorney competence.⁸⁴

⁸⁰ ASS’N OF THE BAR OF THE CITY OF N.Y., REPORT ON ADMISSION TO THE BAR IN NEW YORK IN THE TWENTY FIRST CENTURY: A BLUEPRINT FOR REFORM 467 (1992).

⁸¹ *See id.* at 470.

⁸² *Id.*

⁸³ *See id.* at 467.

⁸⁴ *Id.* at 468. The recommendations of the report resulted only in the addition of one performance question, while subsequent data collection in New York confirmed the persistence of the exam’s disparate racial impact; *see* MICHAEL KANE ET AL., NAT’L CONF. OF BAR EXAMINERS: IMPACT OF THE INCREASE IN THE PASSING SCORE ON THE NEW YORK BAR EXAM 6 (Oct. 4, 2006), http://www.nybarexam.org/press/nyrep_feb06.pdf.

[T]he differences in pass rates among the different racial/ethnic groups are quite large, with the Caucasian/White group having the highest pass rates (about 88% for a passing score of 660 and about 85% for a passing score of 675), and the Black/African American group having the lowest passing rates (about 58% for a passing score of 660 and about 50% for a passing score of 675).

As a result of the 1992 report, the New York Court of Appeals commissioned a study of the bar exam by a team of psychometricians and testing professionals who were charged to consider, among other issues, bar exam content validity (the extent to which the test measures all aspects of lawyer competence),⁸⁵ construct validity (the degree to which the test measures what it claims to be measuring),⁸⁶ and race and gender performance.⁸⁷ To help determine content validity, the Court of Appeals appointed panels of New York lawyers to consider three questions: what areas of law do experienced lawyers think should be tested and with what emphasis; what does the bar exam test besides knowledge of the law; and what other competencies should be tested.⁸⁸ Defining the skills necessary for competent practice as those whose absence would be apt to harm a client, the study identified the key skills⁸⁹ and knowledge⁹⁰ necessary for the competent practice of law. Given the breadth of those skills, the study concluded that the exam could be “advantageously expanded” because it is “far from a perfect sampling of all important lawyering skills.”⁹¹

With respect to construct validity, the report concluded that the bar exam is both valid and reliable (that is, it consistently measures what it claims to be measuring—generalized legal knowledge and legal

Id. at 6.

⁸⁵ JASON MILLMAN ET AL., AN EVALUATION OF THE NEW YORK STATE BAR EXAMINATION 3-1 (May 1993) (defining content validity as the extent to which the test measures all aspects of lawyer competence).

⁸⁶ *Id.* at 9-1.

⁸⁷ *Id.* at 10-1.

⁸⁸ *Id.* at 3-1.

⁸⁹ *Id.* at 3-13. Legal analysis and reasoning; legal research; factual investigation and analysis; problem solving and case planning; written communication; personal qualities of integrity, diligence, timeliness and sound ethical awareness; interpersonal tasks including interviewing, negotiating and counseling; and oral communication and advocacy in the motion and appellate contexts. *Id.*

⁹⁰ *Id.* Knowledge of some core body of doctrinal and procedural law, knowledge of ethical mandates, and knowledge of basic concepts underlying the common law and constitutional law and statutory interpretation. *Id.*

⁹¹ *Id.* at 3-15. The skills and knowledge identified by the study are detailed in the previous two footnotes. This work is consistent with the more recent work done by Marjorie Schultz and Sheldon Zedeck who catalogued the essential competencies by interviewing practicing lawyers and asking them to identify what characteristics they need and value most in their associates. Marjorie M. Shultz & Sheldon Zedeck, *supra* note 1, at 620.

reasoning),⁹² but raised questions about the exam's "speededness."⁹³ A "speeded" exam is one for which the results are dependent on the rate at which the work is performed as well as on the correctness of the response.⁹⁴ Exams differ in their degree of speededness and the impact of speededness on various test-takers.⁹⁵ The report cited evidence that the bar exam is a speeded exam and that doubling the time allowed for the MBE would likely produce a 30 point increase on the New York exam results.⁹⁶ Notably, the report concluded that "speed in reading fact patterns, selecting answers, and writing essay responses [is] not the kind of speed needed to be a competent lawyer."⁹⁷

The report also confirmed that there was a significant gap in passage rates based on race/ethnicity.⁹⁸ While it found no evidence of

⁹² MILLMAN ET AL., *supra* note 85, at 9-17.

⁹³ *See id.* at 9-6.

⁹⁴ For a definition of "speeded," see Kristin Booth Glen, *Thinking Out of the Bar Exam Box: A Proposal to "MacCrate" Entry to the Profession*, 23 PACE L. REV. 3507 (2003) ("speeded," i.e., requiring speed for success).

⁹⁵ *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, 991 fn 65 (2004) (noting that on the LSAT, speededness is currently measured by calculating the proportion of test takers who do not reach each item on the test).

⁹⁶ MILLMAN ET AL., *supra* note 85, at 9-8.

⁹⁷ *Id.*

⁹⁸ *Id.* at 10-4. For the July 1992 bar exam, the passing rate for Asian Americans was 53%; for African Americans 37.4%; for Hispanics 48.6%; and for Whites 81.6%. This disparity is consistent with the findings of disparate impact reported nationally by the LSAC. *See* WIGHTMAN, *supra* note 76 at 27. Additionally, African American law graduates are five times more likely than white graduates to fail the bar examination on the first taking, and while the eventual pass rates were better, the disparity nevertheless persisted. *Id.* at 27 Table 6 (noting that 38.6% of first time black bar examinees fail compared to 8.07% of whites). For instance, studies performed for New York by the National Conference of Bar Examiners show that the differential in performance by different ethnic groups has persisted even though the breadth of the difference has decreased over time. *See* FORDHAM URBAN LAW J., REPORT OF THE NEW YORK STATE JUDICIAL COMMISSION ON MINORITIES 268 (1991) *available at* <http://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=1359&context=ulj> (noting that from 1985–1988, black bar examinees had a first time pass rate of 31% compared to a 73.1% rate for whites). The black/white differential is still 72.3% vs. 92.1%. The racial disparity in pass rates was further confirmed in 2006. *See* KANE, *supra* note 84, at 88.

facial racial bias, the report acknowledged that more subtle bias might play a role.⁹⁹

Since 1993, the New York bar has conducted several additional inquiries and repeatedly raised similar concerns. In 1996, the Professional Education Project, in a study commissioned by Court of Appeals Chief Judge Judith Kaye, recommended reducing the number of subjects tested and developing alternative testing techniques to permit assessment of a wider range of skills.¹⁰⁰ In 2002, the Committees of Legal Education and Admissions to the Bar of the State Bar Association and the Bar of the City of New York issued a joint report criticizing the bar exam for testing only a few of the core competencies required to practice law and citing the national longitudinal study that showed a significant and serious disparate racial impact.¹⁰¹ In 2005, the President of the New York State Bar Association (NYSBA) created the “Special Committee to Study the Bar Exam and other Means of Measuring Lawyer Competence.” After five years of study and debate, the Committee produced the Kenney Report in 2010. The Report recommended streamlining the exam to test more realistically for the knowledge of critical legal rules that should be memorized, reducing the amount of rote memorization, and testing other skills. In 2012, the New York State Bar Association Committee on Legal Education and Admissions to the Bar responded to a call by the NYSBA President to review the Kenney Report and issued a report recommending several proposals to link licensing to more of the skills required in the profession.¹⁰²

⁹⁹ See MILLMAN ET AL., *supra* note 85, at 10-15.

¹⁰⁰ PROF'L EDUC. PROJECT, LEGAL EDUCATION AND PROFESSIONAL DEVELOPMENT IN NEW YORK STATE (1996).

¹⁰¹ See COMMS. ON LEGAL EDUC. & ADMISSION TO THE BAR OF THE ASS'N OF THE BAR OF THE CITY OF N.Y. & THE N.Y. STATE BAR ASS'N, PUBLIC SERVICE ALTERNATIVE BAR EXAM (June 14, 2002), *available at* <http://www.nysba.org/WorkArea/DownloadAsset.aspx?id=26667>

¹⁰² N.Y. STATE BAR ASS'N COMM. ON LEGAL EDUCATION AND ADMISSIONS TO THE BAR, RECOMMENDATIONS FOR IMPLEMENTATION OF THE REPORT OF THE SPECIAL COMMITTEE TO STUDY THE BAR EXAMINATION AND OTHER MEANS OF MEASURING LAWYER COMPETENCE (Feb. 12, 2013). The Report recommended streamlining the exam and assisting the Board of Law Examiners in determining essential content, creating a Practice Readiness Evaluation Program (PREP) that would award points on the bar exam for successful completion of a duly certified clinical course in law school, creation of a pilot project to test a Public Service Alternative to the Bar Exam, studying whether speededness is a

Most recently, in the fall of 2013, the New York City Bar Association issued a Task Force Report that addressed a range of critical issues facing the profession today.¹⁰³ Among its findings, the Task Force reported that “innovation in new lawyer preparation and practice is inhibited by a number of structural impediments that must be removed, through [inter alia] reform of bar exams to permit greater mobility and to focus on the skills needed for success, rather than rote memorization of legal concepts.”¹⁰⁴ The Report acknowledged that bar exams require students to “learn the breadth of the law” but noted:

[B]ar exams are also subject to significant criticism. In particular, critics argue that they are antiquated and fail to test the relevant skills needed to be a lawyer in the twenty-first century. First, the exams ask questions that can easily be answered through legal research. Second, the exams test an applicant’s memory about information that will quickly be forgotten after the exam. Third, most lawyers specialize in their practices, rendering the majority of the information learned for a bar exam irrelevant. (Criminal lawyers have little use for the intricacies of state commercial paper law; corporate deal makers do not need to know state-specific civil procedure.) Finally, and perhaps most importantly in an age requiring graduates to be practice ready, bar exams test few lawyering skills.¹⁰⁵

In its Findings and Recommendations, the Report called on the New York Board of Law Examiners and the New York Court of Appeals to consider revising the content of the bar exam “to test both substantive law and legal practice skills such as complex problem-solving, project management, and exercising professional judgment. . . eliminate[ing] and replace[ing], at least in part, certain areas of the MBE and MPRE in favor of more innovative practice-oriented testing,” and moving

necessary lawyering skill and assessing whether it is producing disparate impacts on women and minorities, and studying the New Hampshire Daniel Webster Program, a two year performance-based bar exam that takes place within law schools, to see whether it could be replicated in New York. The alternatives noted here are discussed in this Article *infra* at notes 159–176.

¹⁰³ NEW YORK CITY BAR, DEVELOPING LEGAL CAREERS AND DELIVERING JUSTICE IN THE 21ST CENTURY, NEW YORK CITY BAR ASSOCIATION TASK FORCE ON NEW LAWYERS IN A CHANGING PROFESSION (Fall 2013).

¹⁰⁴ *Id.* at 5.

¹⁰⁵ *Id.* at 78.

away from “the bar examination’s focus on rote memorization [which] does not benefit any identifiable constituency.”¹⁰⁶

As illustrated in the more than 20 years of review and criticism by the organized bar of New York, concerns about using the bar exam as an all-or-nothing gateway to the legal profession are not new, and are not confined to the members of the Society of American Law Teachers and a small group of academics. Nor do the critiques emanate from a “race above all” ideology. The New York State organized bar’s repeated calls for reform reveal that SALT’s critique of the bar exam is shared by practicing lawyers who understand the importance of ensuring that licensing truly acts as an accurate measure of competence to practice law. What is perhaps most remarkable about the history described here is the marginal difference the New York bar calls for reform have made in the bar exam itself. The most significant change resulting from the many reports, task forces, and recommendations was the addition of one Multistate Performance Test question to the New York bar exam. While responding to one aspect of the critiques, that addition does little to address the broad range of criticism leveled at the bar exam, leaving it still an inadequate measure of lawyering competence and a racially disparate gateway exam that still cannot be justified. In the sections below, we explore the evidence that leads us, and prominent members of the New York bar, to reach that conclusion.

B. What’s Wrong with the Bar Exam?

Even granting that the bar exam tests only a subset of the skills necessary for law practice, are those skills so fundamental, and does the bar test those skills sufficiently well, to warrant using it as a threshold exam? If applicants cannot pass the bar exam, is it worth testing any other skills they might have? Professor Subotnik asks: “If a job requires two skill sets, say jumping and skipping, and if only jumping can be successfully tested, does equity really require that the measurable skill be left untested?”¹⁰⁷ The analogy is incomplete, however. One has to ask whether the test accurately tests jumping, and if jumping—as defined in and tested by the exam—is the same skill of jumping required in the job. It is both what and how you test that matters. While the skills the bar exam purports to test—reading, issue

¹⁰⁶ *Id.* at 85.

¹⁰⁷ Subotnik, *supra* note 2, at 343.

spotting, legal analysis and writing—are clearly fundamental lawyering skills, it is not self-evident that the manner in which these skills are tested indicates whether bar applicants possess these requisite lawyering skills and can apply them in the manner lawyers use them. Contrary to Professor Subotnik’s argument, neither SALT, nor we, argue that states should license incompetent lawyers; the question is whether the bar exam, as presently constituted, is the appropriate “gateway to the profession” test.

1. Bar Exam Format

The bar exam is a timed test, created or adopted by each state’s bar examiners, and generally consists of a multiple choice section, essay questions, and one or two “performance” questions.¹⁰⁸ In most states, the bar exam lasts two days and tests knowledge of majority and minority jurisdiction legal rules, and exceptions to those rules, from material covered in eighteen or more one- or two-semester doctrinal law courses. All jurisdictions except Louisiana and Puerto Rico administer the multiple choice section of the bar exam (the “Multi-State Bar Exam”) written by the National Conference of Bar Examiners (NCBE) and containing 200 multiple choice questions covering seven doctrinal areas.¹⁰⁹ The NCBE also offers essay questions that test majority and minority rules; a state may grade the responses based on its own law or may write and use its own state-focused essay questions.¹¹⁰ Whichever method is used, the essay portion of the exam generally has questions drawn from about eighteen

¹⁰⁸ See *About NCBE Exams*, NCBEX.ORG, <http://www.ncbex.org/about-ncbe-exams/> (last visited Aug. 10, 2014) (describing the various components of the bar exam). Currently, over 35 states include at least one multi-state “performance” question on their exam. See *MPT FAQs*, NCBEX.ORG, <http://www.ncbex.org/about-ncbe-exams/mpt/mpt-faqs/> (last visited Aug. 10, 2014). States also determine the state’s passing score and the weight given to each section of their bar exam. See Gary S. Rosin, *Unpacking the Bar: Of Cut Scores and Competence*, 32 J. LEGAL PROF., 67, 68 (2008) (discussing various states’ passing scores and grading methods).

¹⁰⁹ *About NCBE Exams*, NCBEX.ORG, <http://www.ncbex.org/about-ncbe-exams/mbe/> (last visited Aug. 10, 2014). The current MPT portion of the bar exam tests: constitutional law, contracts, criminal law, criminal procedure, evidence, real property and torts. In 2015, it will add questions on civil procedure. *Multistate Bar Examination News*, NCBEX.ORG, <http://www.ncbex.org/home/multistate-bar-examination-news/> (Sep., 2013).

¹¹⁰ See NCBE, A COMPREHENSIVE GUIDE TO BAR ADMISSION REQUIREMENTS (2014) (describing the components of each state’s bar exam).

doctrinal subject areas.¹¹¹ The third section of the test, the performance test, requires examinees to review a packet of factual and legal information and draft a legal document in 90 minutes.¹¹² As of January 1, 2014, fourteen states had adopted the Uniform Bar Exam [UBE], which tests majority/minority rules using multiple choice, essay and “performance” questions.¹¹³

2. Focus on Memorization

With the exception of the Multi-State Performance Test questions, the bar exam is a closed book test. The volume of material from which test questions are drawn covers thousands of pages in bar preparation materials. Professor Subotnik argues that passing the bar exam is a measure of bar applicants’ ability to learn the law¹¹⁴ and that the volume of material to be memorized is laudable because it indicates command of foundational rules.¹¹⁵ To support his premise that those who question the bar exam denigrate learning, Professor Subotnik argues that memorization is useful because the opposite of memorization is ignorance.¹¹⁶ This follows Professor Subotnik’s pattern of relying on logical fallacies, in this case equating

¹¹¹ *Overview of the MEE*, NCBEX.ORG, <http://www.ncbex.org/about-ncbe-exams/mee/overview-of-the-mee/> (last visited Aug. 10, 2014). The exact number of subjects tested varies state by state. In addition to the areas tested by the MBE, the multi-state essay exam tests business associations (agency and partnership; corporations and limited liability companies), conflict of laws, constitutional law, contracts, criminal law and procedure, evidence, family law, federal civil procedure, real property, torts, trusts and estates (decedents’ estates; trusts and future interests), and uniform commercial code (negotiable instruments and bank deposits and collections; secured transactions). Some questions may include issues in more than one area of law.

¹¹² *Id.*

¹¹³ *The Uniform Bar Examination*, NCBEX.ORG, <http://www.ncbex.org/about-ncbe-exams/ube/> (last visited Aug. 10, 2014) (listing of states that adopted the UBE, including a description of the UBE’s content and how it is graded). States determine their own passing scores. *Id.* New York’s Chief Judge recently proposed adopting the UBE effective July 2015. *See* Joel Stashenko, *Court Seeks Comment on Adopting Uniform Bar Exam*, N.Y.L.J., Oct. 7, 2014, at 1 available at <http://www.newyorklawjournal.com/id=1202672451929/Court-System-Seeks-Comment-on-Adopting-Uniform-BarExam?slreturn=20140920120020>.

¹¹⁴ Subotnik, *supra* note 2, at 369.

¹¹⁵ *Id.* at 371.

¹¹⁶ *Id.* at 371–72.

memorization of the law with knowledge of the law, and failure to memorize the law with ignorance of the law.

Memorization does not equate to retained legal knowledge. Nor is it necessary to memorize a vast body of specific rules in order to capably practice law. In fact, to the extent one's working memory is bombarded with too much information, there is a cognitive overload that can result in forgetting information crucial to understanding, and thus less overall learning.¹¹⁷ While memorizing the law is not enough to pass the bar exam,¹¹⁸ test-takers must be able to retrieve from memory the relevant legal rule or exception. This is unlike accessing the law in legal practice, in which lawyers, after they have identified the legal issue, usually must—and should—research the issue to determine the answer. Being a good lawyer is not about knowing answers immediately. Rather, good lawyers know enough to ask the right questions, figure out how to approach the problem and research the law, or know enough to recognize that the question is outside of their expertise and should be referred to a lawyer more well-versed in that area of the law.¹¹⁹ While lawyers must have a baseline knowledge

¹¹⁷ Hillary Burgess, *Deepening the Disclosure Using the Legal Mind's Eye: Lessons from Neuroscience and Psychology that Optimize Law School Learning*, 29 QUINNIPIAC L. REV. 1, 40–41 (2011) (discussing the application of cognitive load theory to law student learning).

¹¹⁸ Darrow-Kleinhaus, *supra* note 79, at 442.

¹¹⁹ See Karen Erger, *Moving Towards Malpractice*, 23 GPSOLO TECHNOLOGY & PRACTICE GUIDE June 2006, available at https://www.americanbar.org/newsletter/publications/gp_solo_magazine_home/gp_solo_magazine_index/2006_jun_malpractice.html.

Being a lawyer ATM, dispensing legal advice at a moment's notice, any time, day or night, is a malpractice trap. . . . It is all too easy to shoot from the hip and end up shooting yourself in the foot. . . . Inevitably, you'll end up making mistakes. There is no shame in saying, 'I need to do some research—can I call you later this afternoon, when I'm back in the office?' Good clients—the type you want to have and keep—will appreciate that you are taking their matters seriously enough to get things right.

Id.; see also, AMERICAN BAR ASSOCIATION, MODEL RULES OF PROF'L CONDUCT R. 1.1 (2013) (setting forth a lawyer's duty of competent representation which requires "the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation").

A lawyer need not necessarily have special training or prior experience to handle legal problems of a type with which the lawyer is unfamiliar. A newly admitted lawyer can be as

of legal rules sufficient to be able to issue-spot, they need not memorize, or be able to memorize, the nuances tested by the bar exam in order to represent clients competently. Testing whether a bar applicant can memorize hundreds of factoids about multiple areas of law does not test a relevant skill. And as many lawyers will attest, they quickly forgot the law they memorized for the bar exam.¹²⁰ Whether the bar exam is the best way to ensure applicants have the necessary baseline understanding of legal doctrine is not established.

When one claims, as does Professor Subotnik, that the bar exam is a measure of ability to learn,¹²¹ the question must be: to learn what? One might conclude that the bar exam best tests the ability to learn how to take the bar exam itself. Literature suggests that taking bar preparation courses and practice tests or academic enrichment courses geared toward bar passage can improve bar pass rates.¹²² Whether those who improve their results on the test are more competent to practice law is much less clear.

competent as a practitioner with long experience. Some important legal skills, such as the analysis of precedent, the evaluation of evidence and legal drafting, are required in all legal problems. Perhaps the most fundamental legal skill consists of determining what kind of legal problems a situation may involve, a skill that necessarily transcends any particular specialized knowledge. A lawyer can provide adequate representation in a wholly novel field through necessary study. Competent representation can also be provided through the association of a lawyer of established competence in the field in question.

Id. at 1.1. cmt. 2.

¹²⁰ See Burgess, *supra* note 117, at 40.

¹²¹ Subotnik, *supra* note 2, at 369.

¹²² See Linda Jellum & Emmeline Paulette Reeves, *Cool Data on a Hot Issue: Empirical Evidence that a Law School Bar Support Program Enhances Bar Performance*, 5 NEV. L. J. 646 (2005); Derek Alphran, Tanya Washington, & Vincent Eagan, *Yes We Can, Pass the Bar. University of the District of Columbia, David A. Clarke School of Law Bar Passage Initiatives and Bar Pass Rates – From the Titanic to the Queen Mary!*, 14 U.D.C. L. REV. 9 (2011); see also, Keith A. Kaufman, V. Holland LaSalle-Ricci, Carol R. Glass, & Diane B. Arnkoff, *Passing the Bar Exam: Psychological, Educational, and Demographic Predictors of Success*, 57 J. LEGAL EDUC. 205, 218 (2007) (finding that “on average, graduates who passed the bar exam on their first try took almost twice as many practice tests as did those who failed” but also finding that the number of practice tests did not affect second-time performance after an initial failure).

3. Impact of Test-Taking Speed

The bar exam, like other standardized tests, measures a variable unrelated to law practice: test-taking speed.¹²³ The multi-state multiple choice exam requires test-takers to complete 100 questions in 180 minutes in the morning and again in the afternoon, so they can spend only an average 1.8 minutes on each question. Multi-state essay questions allocate approximately a half hour to answer each of six essay questions, many of which may be multi-part questions.¹²⁴ The multi-state performance test, a test designed to assess test takers' ability to read and interpret the law in light of a client's problem and to produce a product similar to those produced by lawyers¹²⁵—and thus the part of the test most aligned with the actual work of attorneys—allows test takers 90 minutes to read the provided cases, statutes, and other materials and then prepare a legal document.¹²⁶ All portions of the bar exam thus require examinees to act quickly as they read and digest the material, recall the applicable law, and apply that law to the given test question.

¹²³ Henderson *supra* note 95, at 979.

¹²⁴ *Overview of the MEE*, NCBEX.ORG, <http://www.ncbex.org/about-ncbe-exams/mee/overview-of-the-mee/> (last visited Aug. 10, 2014). (The multi-state essay exam consists of six thirty minute questions). States using the UBE must use the multi-state essay questions. Other states may use the questions. For a listing of which states use the MEE, see *MEE FAQs*, NCBEX.ORG, <http://www.ncbex.org/about-ncbe-exams/mee/mee-faqs/> (last visited Aug. 10, 2014).

¹²⁵ *Overview of the MPT*, NCBEX.ORG, <http://www.ncbex.org/about-ncbe-exams/mpt/overview-of-the-mpt/>. The NCBE says that the multi-state performance test measures the ability to:

- (1) sort detailed factual materials and separate relevant from irrelevant facts;
- (2) analyze statutory, case, and administrative materials for applicable principles of law;
- (3) apply the relevant law to the relevant facts in a manner likely to resolve a client's problem;
- (4) identify and resolve ethical dilemmas, when present;
- (5) communicate effectively in writing; and
- (6) complete a lawyering task within time constraints.

Id.

¹²⁶ *Id.* The NCBE provides two 90-minute questions for each administration of the MPT; states may choose to use one or both of them, but states administering the Uniform Bar Exam must use both. *See id.* (stating jurisdictions that use the UBE must administer two MPT questions and noting that jurisdictions, other than those using the UBE, may choose to use one or both of the MPT questions).

To illustrate the impact of speededness on bar exam performance, we include a closer look at an example of a bar exam question, offered by Professor Suzanne Darrow-Kleinhaus as an illustration that more than memorization is required to succeed on the bar exam.¹²⁷

Peavey was walking peacefully along a public street when he encountered Dorwin, whom he had never seen before. Without provocation or warning, Dorwin picked up a rock and struck Peavey with it. It was later established that Dorwin was mentally ill and suffered recurrent hallucinations.

If Peavey asserts a claim against Dorwin based on battery, which of the following, if supported by evidence, will be Dorwin's best defense?

- A) Dorwin did not understand that his act was wrongful.
- B) Dorwin did not desire to cause harm to Peavey.
- C) Dorwin did not know that he was striking a person.
- D) Dorwin thought Peavey was about to attack him.¹²⁸

Professor Darrow-Kleinhaus notes that test-takers who answer A or B may have “react[ed] to [the] answer instead of applying the elements methodically to the issue.”¹²⁹ Test-takers should recognize, she says, that the intent element of battery is satisfied either when a harmful or offensive contact is intended or when the tortfeasor acts with purpose or with knowledge to a “substantial certainty” that the result will follow.¹³⁰ But that oversimplifies the analytical process. To “methodically” apply the elements, test-takers should engage in the following sequence of thoughts:¹³¹

- The elements of battery are (a) intent to cause a harmful or offensive contact and (b) a harmful or offensive bodily contact results.

¹²⁷ Darrow-Kleinhaus, *supra* note 79, at 447–49. As noted earlier, retrieving relevant rules from one's memory is the necessary first step to successfully answering this question, and we have challenged the value of that requirement. *See supra* text accompanying notes 117–120. Our point here is that speededness is also a problem because the test requires retrieval of memorized legal rules and then a substantial amount of analysis.

¹²⁸ Darrow-Kleinhaus, *supra* note 79, at 447–48.

¹²⁹ *Id.* at 448.

¹³⁰ *Id.*

¹³¹ *See Id.*

- The definition of intent is (a) having the purpose or desire to cause a harmful or offensive bodily contact or (b) knowing with substantial certainty that such a result will occur.
- How do those elements apply to answer A? If Dorwin didn't understand his act was wrongful, he still may have had the requisite intent. Perhaps he thought that Peavey was Satan and he was saving the world. He would not have thought his action was wrongful but he would nevertheless satisfy the definition by having the purpose of causing a harmful or offensive bodily contact or knowing with substantial certainty that such a contact would occur. Answer A is incorrect.
- How do those elements apply to answer B? If he knew that the harm was substantially certain to occur, then even if Dorwin didn't want to cause harm, he would still be liable, so answer B is incorrect.
- Considering answer C, does battery require that you know you are harming a person? If Dorwin did not know he was hitting a person, he probably would not be intending to harm a person or be substantially certain he was harming a person. Thus, C may be correct because if Dorwin did not know Peavey was a person he could not have formed the intent to commit a battery, which is only a tort against a person. Or, as Professor Darrow-Kleinhaus notes, "only. . .C completely negates the intent element: If Dorwin had no idea (no 'knowledge') he was striking a person, then he could not have formed the requisite intent to do the act."¹³²
- Before choosing answer C, the test taker would also have to eliminate answer D. Although self-defense would be a viable defense "if supported by evidence," Professor Darrow-Kleinhaus notes that the test-taker should realize answer D is incorrect because the facts do not indicate that Peavey was about to attack Dorwin and that the facts rule out this defense because "Peavey was walking peacefully" and it was "without provocation or warning [that] Dorwin picked up a rock."¹³³ Thus, before eliminating answer D, the test-taker would likely need to return to the question and re-read the facts to know whether the answer could be correct. The test-taker would also have to know that self-defense requires not only an actual but a reasonable belief in imminent harm, and that Dorwin might have thought he was in such a position as a result of his mental illness but his belief would not be reasonable.

¹³² *Id.* at 448.

¹³³ *Id.*

Professor Darrow-Kleinhaus uses this example to illustrate that “merely memorizing the elements of battery is insufficient because the bar exam requires an analysis of the question followed by an analysis of each possible answer with respect to the legal issue posed above.”¹³⁴ We suggest that the question and resulting analytical process illustrates the role test-taking speed plays in this gateway exam. Reading the question closely and engaging in the analytical process described above must happen in approximately one minute and 48 seconds. This same process must be undertaken 200 times for questions drawn from seven doctrinal areas.

That the bar exam is time-pressured cannot be denied. But is the ability to successfully take such a timed exam important to establishing competence as a lawyer? Professor Subotnik argues it is because lawyers often must meet deadlines. He suggests, without any empirical basis, that test-taking speed is related to lawyer efficiency.¹³⁵ While it is true that lawyers must efficiently organize their time, the bar exam does not purport to test one’s ability to do that and there is no evidence that test-taking speed is related to lawyering skill. In fact, at least one study shows that “time management ability and its components . . . failed to predict bar exam passage.”¹³⁶ As noted by Professor William Henderson in his study of the role of speededness on the LSAT and in law school exams:

Time is certainly relevant in the legal profession. Lawyers bill by the hour. They are also occasionally pressed by clients to provide immediate legal advice over the phone without the benefit of any research or reflection. An objection to an evidentiary issue cannot be the subject of an appeal unless it has been timely raised before the trial court. Similarly, appellate judges pride themselves on raising novel and unexpected issues during oral argument.¹³⁷

But as Professor Henderson further notes, the need for efficiency in some (but not all) aspects of effective lawyering does not answer “[t]he more difficult analytical question” whether the time facility required on the LSAT or law school exams (or, we add, the bar exam) is “an accurate metric for these widely divergent concepts of efficiency

¹³⁴ *Id.* at 449.

¹³⁵ Subotnik, *supra* note 2, at 372.

¹³⁶ Kaufman et al., *supra* note 122, at 217.

¹³⁷ Henderson, *supra* note 95, at 1035 (footnotes omitted).

and speed.”¹³⁸ Based on his study of student performance on the LSAT and on in-class and much longer take-home exams, Professor Henderson concluded that time-pressured law school exams are not an effective measure of efficient writing ability or high performance in oral advocacy.¹³⁹ The same may be said for time-pressured bar exams. As noted earlier, a New York bar study concluded that doubling the time allowed for the MBE would likely produce a thirty point increase on the New York exam results¹⁴⁰ and that “speed in reading fact patterns, selecting answers, and writing essay responses is not the kind of speed needed to be a competent lawyer.”¹⁴¹

4. Bar Exam Results and Law School Grading

Professor Subotnik suggests that the correlation between law school grade point average [LGPA] and bar pass rate shows the bar exam is a measure of ability to learn and apply the law.¹⁴² A correlation between LGPA and bar pass rates is not surprising because many law school exams, especially in the first year, test students on the same skills as the bar exam and in similar ways—with multiple-choice and essay questions relying on “speededness”—so the limitations of the bar exam as an accurate and comprehensive test of lawyer competence hold as well for those correlated exams. Neither indicates the ability to succeed in law practice, as further explored below. The question should not be whether LGPA relates to bar pass scores but rather whether the assessment methodologies used by both legal educators and bar examiners adequately capture the skills lawyers need as those skills are utilized in practice. As law schools expand the use of experiential coursework and non-speeded evaluation, the correlation between the two sets of exams may decrease. However, the current correlation is not by itself a reason to give more credence to the bar results.

5. Bar Exam Results and Lawyering Competence

Even if the bar exam (and law school exams) adequately test the admittedly important skill of learning legal rules and applying them to

¹³⁸ *Id.*

¹³⁹ *Id.* at 1036–38.

¹⁴⁰ MILLMAN ET AL., *supra* note 85, at 9–8 & n. 11.

¹⁴¹ *Id.*

¹⁴² Subotnik, *supra* note 2, at 379–80.

new facts, it is far from clear that such exams test the ability to learn and apply the law in the context of client representation. As Professor Cecil Hunt notes, “while the bar examination may be an excellent test of the ability to study law competently, it does not necessarily indicate the ability to practice law competently.”¹⁴³ Although we know of no large-scale study correlating bar exam scores or law school grades with professional success,¹⁴⁴ a large New York law firm evaluated all the lawyers it hired over a thirty-year period and concluded that, with the exception of the top one to two percent of top law school performers, “there was little to no correlation between law school grades and the work performance of those who attained partnership.”¹⁴⁵

It would not be surprising to conclude that there is little correlation between law school grades and lawyers’ success, because many law school assessments do not measure the wide range of skills lawyers need, including creativity, the ability to work well in teams, listening skills, and common sense and good judgment.¹⁴⁶ One multi-year study found no correlation between law school grades and achievement in a simulation-based negotiations class, in which the final grade was based upon bargaining results from a series of negotiation exercises and student papers analyzing their negotiation-exercise experiences.¹⁴⁷ Examining student grades over thirteen semesters in a class that ranged

¹⁴³ Cecil J. Hunt, II, *Guests in Another’s House: An Analysis of Racially Disparate Bar Performance*, 23 FLA. ST. U. L. REV. 721, 767 (1996).

¹⁴⁴ Lani Guinier, *Lessons and Challenges of Becoming Gentlemen*, 24 N.Y.U. REV. L. & SOC. CHANGE 1, 12 (1998) [hereinafter Guinier, *Lessons and Challenges*]; see also, David B. Wilkins, *A Systematic Response to Systemic Disadvantage, A Response to Sander*, 57 STAN. L. REV. 1915, 1925 (2005) (“studies that attempt to understand the causes of long-term career success have not documented a robust correlation between grades and other entry-level credentials and long-term career success”).

¹⁴⁵ Guinier, *Lessons and Challenges*, *supra* note 144 at 12; see William C. Kidder, *The Bar Examination and The Dream Deferred: A Critical Analysis of the MBE, Social Closure, and Racial and Ethnic Stratification*, 29 L. & SOC. INQUIRY 547, 580 (2004) (discussing FTC study which found no correlation between the agency’s lawyers’ accomplishments and their bar exam performance or LSAT scores or LGPA) [hereinafter Kidder, *The Bar Exam*].

¹⁴⁶ See, e.g., Shultz & Zedeck, *supra* note 1, at 630, Table 1 (listing lawyering effectiveness factors).

¹⁴⁷ Charles Craver, *The Impact of Student GPAs And A Pass/Fail Option On Clinical Negotiation Course Performance*, 15 OHIO ST. J. ON DIS. RES. 373 (2000).

from 40 to 62 students per semester, the instructor found that student grades on the exercises did not correlate with the students' overall LGPAs. He suggests that this lack of correlation stems from the fact that traditional law school courses test abstract reasoning skills while the negotiation exercises measure emotional intelligence skills (e.g., the ability "to 'read' other people").¹⁴⁸ The course papers also did not assess traditional legal analysis; instead, they required students to engage in self-reflection about their bargaining interactions and how those interactions related to negotiation theory. This study confirms the intuition that the various kinds of intelligence needed by successful lawyers are not distributed in the same proportion in law students, bar applicants, and lawyers, and that as schools begin to integrate more experiential learning into the curriculum and assess skills beyond those related to abstract reasoning, the correlation between LGPA and bar pass may decline.

6. What the Bar Exam Does Not Test

A set of recent empirical studies, one of them commissioned by the National Conference of Bar Examiners, explore the work performed by newly licensed lawyers and the skills necessary for effective lawyering; both studies indicate the bar exam tests only a small portion of the skills new lawyers need.¹⁴⁹ In its comprehensive study of the work performed by newly licensed lawyers, the NCBE sought to identify the tasks, knowledge domains, skills, and abilities significant to newly licensed lawyers to "provide a job related and valid basis for the development of licensing examinations offered by the NCBE."¹⁵⁰ The study identified 329 tasks performed by a wide range of newly licensed lawyers with various specialties including: 43 "general tasks," 86 "knowledge domains," and 36 general "skills and abilities."¹⁵¹ The

¹⁴⁸ *Id.* at 383.

¹⁴⁹ See Shultz & Zedeck, *supra* note 1, at 623 (discussing the empirically derived list of lawyering effectiveness factors); Steven Nettles & James Hellrung, *A Study of the Newly Licensed Lawyer*, available at http://www.ncbex.org/assets/media_files/Research/AMP-Final-2012-NCBE-Newly-Licensed-Lawyer-JAR.pdf [hereinafter *NCBE study*] (setting out the numerous skills new lawyers need in order to perform their jobs).

¹⁵⁰ *NCBE study*, *supra* note 149, at 1.

¹⁵¹ *Id.* at 289–98. Those tasks include aspects of managing the attorney-client relationship and caseload, communicating with clients and others, handling research and investigation, and analyzing and resolving client matters. The skills and abilities fell into similar categories. *Id.*

list reveals a multitude of lawyering tasks and skills the existing bar exam does not attempt to measure.

The failure of the bar exam to test most of the skills necessary for practicing attorneys undermines the rationale for according it gateway status to the profession and adds to the justification for exploring additional and alternative means to test applicants' entry-level competence. While Professor Subotnik suggests that critiques of the bar exam equate to anti-intellectualism,¹⁵² we suggest the opposite: the anti-intellectuals are those who rigidly insist the current exam is the best we can do and are unwilling to explore alternatives to what even its advocates admit is an imperfect exam.

C. The Bar Exam Constrains Change in Legal Education

Spurred by the economic downturn, by reports from the Carnegie Foundation¹⁵³ and the Clinical Legal Education Association,¹⁵⁴ and by criticism from the practicing bar and their own self-evaluations, law schools increasingly recognize that pure doctrinal teaching ill-equips students to represent clients.¹⁵⁵ In response to these critiques, schools across the country are integrating experiential learning into their curricula and are expanding students' experiential learning opportunities.¹⁵⁶ Instructors in many, if not all doctrinal areas, have begun integrating more experiential exercises into doctrinal courses.¹⁵⁷

¹⁵² Subotnik, *supra* note 2, at 369–70 (arguing that SALT does not care if future lawyers are learned in the law); *id.* at 398 (suggesting that those who critique tests actually critique knowledge and “devalue intellectual achievement”).

¹⁵³ WILLIAM M. SULLIVAN ET AL., EDUCATING LAWYERS: PREPARATION FOR THE PROFESSION OF LAW (2007).

¹⁵⁴ Roy Stuckey et al., *Best Practices for Legal Education: A Vision and a Roadmap*, CLINICAL LEGAL EDUC. ASS'N. (2007), available at <http://www.cleaweb.org/Resources/Documents/best-practices-full.pdf>.

¹⁵⁵ A. Benjamin Spencer, *The Law School Critique in Historical Perspective*, 69 Wash. & Lee L. Rev. 1949, 2001-15 (2012) (discussing decades of critiques about legal education's failure to adequately prepare law students for law practice).

¹⁵⁶ As of December 2013, over 30 law schools had appointed deans or directors of experiential learning. E-mail from Alli Gerkman, Dir., Educating Tomorrow's Lawyers (January 14, 2014, 16:13 EST) (on file with author Andrea Curcio). Just a few years ago, no such position existed at most law schools. A web search in October 2014 revealed dozens of schools advertising their experiential programs.

¹⁵⁷ See, e.g., Bradley T. Borden, *Using the Client-File Method to Teach Transactional Law*, 17 CHAPMAN L. REV. 101 (2013); Paula Schaefer, *Injecting*

Students are strongly encouraged to explore clinics and externships and other experiential learning opportunities. A recently adopted ABA Accreditation Standard requires all law students to take at least six experiential learning credits¹⁵⁸ and California's State Bar Board of Trustees has approved a competency training proposal for new lawyers that includes, among other things, 15 units of practice-based, experiential course work or an equivalent apprenticeship during law school.¹⁵⁹

As schools consider such changes, they will face inevitable tension between offering experiential learning courses and bar-exam-focused courses. While one study indicates that taking "bar courses" has little discernible impact on bar pass rates,¹⁶⁰ both law students and law schools make curricular choices based on the assumption that it does.¹⁶¹ Especially in economic downturns such as the current one,

Law Student Drama Into the Classroom: Transforming an E-Discovery Class (or Any Law School Class) with a Complex, Student-Generated Simulation, 12 NEV. L. J. 130 (2011); Roberto Corrada, *A Simulation of Union Organizing in a Labor Law Class*, 46 J. Leg. Educ. 445 (1996); Anne M. Tucker, *LLCs by Design*, 71 WASH. & LEE L. REV. 525 (2014).

¹⁵⁸ Standard 303(a)(3), 2014-2015 ABA STANDARDS AND RULES OF PROCEDURE FOR APPROVAL OF LAW SCHOOLS, available at http://www.americanbar.org/groups/legal_education/resources/standards.html.

¹⁵⁹ Amy Yarborough, *Board Asks for Next Steps in Competency Training Plan*, CAL. ST. B.J. (Nov. 2013) available at <http://www.calbarjournal.com/November2013/TopHeadlines/TH4.aspx>.

¹⁶⁰ Douglas K. Rush & Hisako Matsuo, *Does Law School Curriculum Affect Bar Examination Passage? An Empirical Analysis of Factors Related to Bar Examination Passage During the Years 2001 Through 2006 at a Midwestern Law School*, 57 J. LEG. EDUC. 224, 234-35 (2007) (discussing how upper level elective choices had no correlation to bar pass rates for most students at St. Louis University School of Law).

¹⁶¹ See, e.g., Donald H. Zeigler et al., *Curriculum Design and Bar Passage: New York Law School's Experience*, 59 J. LEGAL EDUC. 393 (2010) (discussing how changes to NYLS's curriculum, including the requirement that students in the bottom quartile of the class take a wide array of courses tested on the bar exam, have improved NYLS' bar passage rates). Other schools have required struggling students to take examination subject matter courses. Rush & Matsuo, *supra* note 160, at 227-28; see also ABA SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, REPORT OF THE TASK FORCE ON LAW SCHOOLS AND THE PROFESSION: NARROWING THE GAP 278 (1992) (commonly known as "The MacCrate Report") (noting that the bar exam influences law schools to develop curricula that overemphasize courses covered by the exam and that the exam influences law students to choose doctrinal courses in areas

where should schools spend limited resources? Does it make sense for schools to focus on teaching more doctrine tested via multiple choice and essay questions, or is it more beneficial to expose students to experiential learning opportunities in which doctrine is integrated into real world experiences? Should the bar exam be the “tail that wags the dog”? These questions require thoughtful debate—something not possible if one simply insists that questioning the existing exam equates to anti-intellectualism.

IV. ALTERNATIVES TO THE EXISTING BAR EXAM

In the circumstances of *Ricci*, as in other instances raising disparate impact concerns, a key question is whether valid but less discriminatory exams exist. The discriminatory impact of the bar exam should lead us to investigate the possibility and viability of less discriminatory alternative methods of testing attorney competence. Instead of supporting the consideration and development of alternatives, Professor Subotnik suggests that we cannot measure skills beyond those measured by the existing bar exam.¹⁶² To counter that argument, this Article reviews several alternative measures currently used, proposed, or under development.

Although these particular innovations may or may not be the solution to the problematic reliance on a limited exam that has racially disparate outcomes, the critical point is that alternatives are possible and, if there is general agreement that skills beyond those tested on the bar exam matter, there will be a concerted effort to develop ways to test those skills. If we decide that skills should or must be assessed in more job-related ways, educators and testing companies will develop better tests to assess more skills. This has already occurred in the medical licensing field under the pressure to ensure that new doctors

tested by the exam); Byron D. Cooper, *The Bar Exam and Law Schools*, 80 MICH. B.J. 72, 73 (2001) (noting that some Michigan law schools saw substantial increases in students enrolling in no-fault automobile insurance and worker’s compensation classes when those subjects were added to the Michigan bar exam; further noting that an informal survey of Michigan property law professors found the majority of professors took “the bar exam into consideration in deciding which sections of the required casebook should be covered in the course.”)

¹⁶² Subotnik, *supra* note 2, at 369 (arguing the bar exam is a meaningful exam); *id.* at 373–78 (critiquing proposed alternatives to the exam).

have more than book-learning about medicine.¹⁶³ As already noted, alternative assessments of firefighters through simulations of real-world situations were available through assessment centers,¹⁶⁴ though the *Ricci* court did not compel their use.¹⁶⁵ Those alternative assessments had a much smaller disparate impact than the contested paper and pencil tests. If individuals developing firefighter exams can create viable alternatives with higher construct validity and less discriminatory impact,¹⁶⁶ we believe that the same result can be achieved by those developing law licensing exams. The disparate impact of current tests is a reason to explore those alternatives—not to reach a particular racially-balanced result, as Professor Subotnik contends, but to reach a result that does a better job at measuring competence.

A. The New Hampshire Model

The Daniel Webster Scholars Honors Program (DWS), launched in 2005 at Franklin Pierce Law School, now the University of New Hampshire School of Law, is designed to produce client-ready lawyers by providing a two year practice-based, client-oriented education in the second and third years of law school. The program educates students, but also operates as an alternative bar exam, as students who successfully complete the program are certified by the Board of Law Examiners and are admitted to the New Hampshire bar upon graduation.¹⁶⁷

The DWS was originally conceived by Chief Justice Linda Dalianis of the New Hampshire Supreme Court, whose many years on

¹⁶³ Jayne W. Barnard & Mark Greenspan, *Incremental Bar Admission: Lessons From the Medical Profession*, 53 J. LEGAL EDUC. 340, 344–45 (2003). The same development has occurred in the field of general cognitive testing; the disparate impact of general intelligence tests like the SAT has led researchers to explore broader conceptions of intelligence and to create tests to assess them. See *infra* Parts V.C, VI.A (discussing tests that can be used to supplement cognitive-based testing for admissions and employment decisions).

¹⁶⁴ *Ricci v. DiStefano*, 557 U.S. 557, 559 (2009).

¹⁶⁵ *Id.* at 609 (Ginsberg, J., dissenting) (citing to better tests used by many departments).

¹⁶⁶ IOP Brief, *supra* note 45 at 11; *Ricci*, 557 U.S. at 558.

¹⁶⁷ See John Burwell Garvey, *Making Law Students Client-Ready – The Daniel Webster Scholar Honors Program: A Performance-Based Variant of the Bar Exam*, N.Y. ST. B.A. J. 46 (Sept. 2013). Students must also satisfy the character and fitness requirement and pass the MPRE. N.H. SUP. CT R. 42, Part XII(a).

the bench convinced her that too many graduates of ABA-accredited law schools lacked the skills and knowledge necessary to practice law effectively. To develop a licensing mechanism that more closely evaluates the knowledge, skills, and values required for effective lawyering, Justice Dalianis created a working group of bar examiners, judges, lawyers, and academics to develop an alternative licensing program—a better bar exam—that would incorporate the “MacCrate” skills¹⁶⁸ and assess students on whether they were truly practice-ready. Her goal was “to make lawyers better.”¹⁶⁹ The Committee persuaded the New Hampshire Supreme Court to amend its bar admission rule to authorize a performance-based variant of the bar exam that “[consisted] of rigorous, repeated and comprehensive evaluation of legal skills and abilities.”¹⁷⁰

Participants in the program must maintain a high grade point average and complete simulation courses specially designed for the DWS program, as well as meet all other law school requirements for graduation. All DWS students are required to engage in pro bono work after receiving training as advocates for victims of domestic violence. Students must also take six credits of externship and/or clinical courses and, as their capstone course, Advanced Problem Solving and Client Counseling, a course “[t]hat integrates and builds upon the skills students have already learned through the program. This course takes them to the next level, particularly emphasizing fact gathering (including witness interviewing), legal analysis, problem solving, and client counseling.”¹⁷¹

Assessment is “an integral part of the DWS program” and students receive “nearly continuous feedback.”¹⁷² The courses and the rubrics used to evaluate performance track the fundamental lawyering skills identified in the MacCrate Report as well as the factors identified in the Schultz/Zedeck study of effective lawyering, discussed below.¹⁷³

¹⁶⁸ See Robert MacCrate, *Yesterday, Today and Tomorrow, Building The Continuum of Legal Education and Professional Development*, 10 CLINICAL. L. REV. 805, 831 (2004).

¹⁶⁹ Hon. Linda S. Dalianis & Sophie M. Sparrow, *New Hampshire’s Performance-Based Variant of the Bar Examination: The Daniel Webster Scholar Program*, THE BAR EXAMINER, November 2005, at 23, 25.

¹⁷⁰ N.H. SUP. CT. R. 42 at Part XII(a)(1).

¹⁷¹ Garvey, *supra* note 167 at 45.

¹⁷² *Id.* at 46–47.

¹⁷³ *Id.* See *infra* notes 234–37 and accompanying text.

Students develop an extensive portfolio, including videos of the student conducting simulated interviews, negotiations, and components of trial practice. Evaluation and assessment are done by law school faculty and members of the New Hampshire Board of Law Examiners, who review the student portfolios each semester and meet personally with the students each year to evaluate their progress. Students are also assessed through the use of standardized clients, a program modeled after the standardized patient used in medical school assessment. The simulated clients are actors trained to assess a student's skill using written standardized criteria.¹⁷⁴

The success of the DWS program is evident in the feedback from employers, who report that graduates of the DWS program are better prepared for practice and far more client-ready than non-participants.¹⁷⁵ The program's success in meeting its goals is also evident from observations by Lloyd Bond, one of the lead authors of the Carnegie Report:

We can only hope that other state Supreme Courts will seriously consider the Webster Scholar method as an alternative approach to training and licensure. When I studied the program in depth . . . I said it fused instruction, assessment, and practice in such an integrated way that the three became indistinguishable. The Daniel Webster Scholar Program exemplifies the sea change we had in mind.¹⁷⁶

William Sullivan, the lead author of the Carnegie Report and the Founding Director of Educating Tomorrow's Lawyers, agrees. In writing about the need to link a practice-based curriculum to licensing,¹⁷⁷ he notes the need to "move students more effectively across the arc of professional development from novice to competent beginning practitioner and . . . to assess the readiness of such

¹⁷⁴ *Id.* at 48.

¹⁷⁵ Educating Tomorrow's Lawyers recently conducted a rigorous assessment of the DWS program. The results, which are expected to be published in Fall 2014, will play a major role in efforts to enhance the certification function of the bar exam.

¹⁷⁶ Lloyd Bond, Consulting Scholar (retired), The Carnegie Foundation for the Advancement of Teaching, Prepared remarks to the Conference on a Performance-Based Approach to Licensing Lawyers: The New Hampshire "Two-Year Bar Examination" (April 23, 2010) (cited in Garvey, *supra* note 167 at 50).

¹⁷⁷ William M. Sullivan, *Align Preparation and Assessment with Practice: A New Direction for the Bar Examination*, 85 N.Y. St. B. J. 41 (2013).

developing lawyers”¹⁷⁸ and cites the DWS program as doing just that.¹⁷⁹

The New Hampshire Daniel Webster Scholars Program demonstrates that it is possible to measure skills beyond those measured by the existing bar exam and that a performance-based bar exam is a better measure of the competencies required for law practice. The fact that the program requires a substantial commitment of educational resources and therefore may not be adopted by most jurisdictions in its current form does not detract from the conclusion that it *is* possible to design courses and assessments that focus on more than the slim set of skills tested by the current bar exam.

B. Modify the Existing Exam

To address one of the criticisms leveled against the current bar exam—that it depends too much on memorization—the New York bar has suggested testing fewer doctrinal subjects.¹⁸⁰ Proponents of the multiple choice section of the exam argue that a large enough sample of multiple choice questions from a representative range of content is needed to ensure content validity,¹⁸¹ but testing fewer doctrinal areas should provide sufficient scope to test the ability to issue spot and engage in legal analysis. And if legal analysis is the key, not simply memorizing the rules, perhaps consideration should be given to making the exam somewhat “open book” by providing a sourcebook for rules so applicants could focus on applying law rather than memorizing law.

To address a second concern about the current bar exam—that it depends too much on speededness—the number of questions, especially multiple choice questions, could be reduced or the time extended to allow test-takers more time per question, leading to a more relevant measure of knowledge and analysis, rather than of the ability to respond quickly in test-taking circumstances.

Perhaps, as Professor Subotnik acknowledges,¹⁸² the cornerstone of a modified bar exam could be a performance test. A well-designed

¹⁷⁸ *Id.*

¹⁷⁹ *See supra* note 175

¹⁸⁰ *See supra* notes 80–81.

¹⁸¹ *Workshop on the Future of the Legal Course Book*, 33 SEATTLE U. L. REV. 292, 305 (2008).

¹⁸² Subotnik, *supra* note 2, at 375.

performance test is a better way of testing lawyering skills than multiple choice and essay questions, which require memorization of large quantities of doctrine. In its current incarnation, however, performance testing is still a highly speeded endeavor. Students must read and digest multiple fact-based documents, cases, and statutory provisions, and write a memo, brief, or client advisory letter, within a ninety minute time period.¹⁸³ While the tasks replicate what lawyers do, the testing method could be modified to bear a closer relationship to law practice and avoid undue compression of testing time. Nonetheless, the MPT, as currently embodied in the MPT and a part of the UBE, is a step in the right direction, especially if more time were allowed to analyze materials and compose documents, and if more of the exam were devoted to performance questions.

To the extent the bar exam is modified so that it truly becomes a test about analyzing and applying the law and not merely remembering it, and avoids having test-taking speed be an independent variable affecting exam scores, the exam could be a vastly better measure of who should be permitted to enter law practice. Changes such as these could be explored and implemented without a radical overhaul of the entire system of bar admission.

C. Test More Skills

If the goal of licensing is to ensure entry-level competence for legal practice, licensing decisions should be based on an applicant's demonstration of competence in the necessary set of skills. If less time were taken testing rules in many subject areas, time would be available to test a wider range of skills such as legal research, interviewing, counseling, and negotiation. With the advent of computer-based testing, testing such skills becomes more plausible.¹⁸⁴ Interviewing, counseling, and negotiation skills could be assessed by asking questions based upon video vignettes, or simulations could be developed to test a wide range of lawyering skills using a virtual dialog with a simulated client, opposing counsel, or judge, similar to

¹⁸³ For summaries of the materials and tasks required by the MPT, see *2013 Multi-state Performance Test Information Booklet* at 7–11, available at http://www.ncbex.org/assets/media_files/Information-Booklets/MPTIB2013.pdf.

¹⁸⁴ See Andrea A. Curcio, *A Better Bar: Why and How the Existing Bar Exam Should Change*, 81 NEB. L. REV. 363, 366, 394–96 (2002) (discussing use of computer simulations to assess a wider range of lawyering skills).

what is already done in medical licensing.¹⁸⁵ The legal profession could develop a “standardized client” assessment based upon the medical licensing standardized patient model.¹⁸⁶

Another practical way to assess an applicant’s competence in relevant skills, at least in part, is to rely on student participation in supervised clinic or externship experiences. This idea was originally proposed over ten years ago by Robert MacCrate,¹⁸⁷ who recognized the value that clinical learning adds to law students’ education, and that successful completion of such experiences demonstrates the development of students’ skills, abilities and attitudes that cannot be tested in the existing bar format.¹⁸⁸ “[J]ust as the results of the multi-state performance test are factored into an applicant’s total score,” he suggested, “credit should be given for successfully completed clinical experiences supervised and certified by the faculty of a duly accredited law school.”¹⁸⁹ An additional benefit of offering such credit is to encourage students to take courses in law school best designed to help them become better lawyers.

Consistent with MacCrate’s proposal, in 2012 the New York State Bar Association Committee on Legal Education and Admissions to the Bar endorsed what it called a “Practice Readiness Evaluation Program” that would add points to a traditional bar exam score for those students who successfully completed a bar-certified clinical education experience.¹⁹⁰ While expressing some concerns,¹⁹¹ the

¹⁸⁵ Barnard & Greenspan, *supra* note 163, at 344 (discussing the use of case-based simulations in medical licensing); *see* PRIMUM COMPUTER BASED SIMULATIONS (CSS) FOR LICENSING DOCTORS, <http://www.jisc.ac.uk/media/documents/projects/higherorderskills.pdf> (last visited Aug. 10, 2014) (describing how case-based simulations in medical licensing works with sample problems).

¹⁸⁶ The theoretical underpinnings of the standardized client were discussed approximately a decade ago. *See, e.g.*, Barnard & Greenspan, *supra* note 163, at 345; Lawrence M. Grosberg, *Standardized Clients: A Possible Improvement for the Bar Exam*, 20 GA. ST. U. L. REV. 841, 841 (2004). Work using standardized clients began with law students. *See* Karen Barton et al., *Valuing What Clients Think: Standardized Clients and the Assessment of Communicative Competence*, 13 CLIN. L. REV. 1, 3–5 (2006). The New Hampshire two year bar examination has incorporated standardized client assessments. Garvey, *supra* note 167, at 49.

¹⁸⁷ *See* MacCrate, *supra* note 168, at 831.

¹⁸⁸ *Id.*

¹⁸⁹ *Id.*

¹⁹⁰ *New York State Bar Association Committee on Legal Education and Admission to the Bar, Recommendations for Implementation of the Report of the Committee*

majority of the committee voted to explore the idea, believing that credit for a clinical experience “was a practical and affordable way” to integrate into the bar exam process assessment of student ability to handle real cases and real clients.¹⁹²

Earlier, in 2002, a Joint Report of the Committees on Legal Education and Admission to the Bar of the City of New York and the New York State Bar Association endorsed a Public Service Alternative to the Bar Exam (PSABE), a performance-based bar exam originally conceived by Kristin Booth Glen, then dean of CUNY Law School.¹⁹³ The PSABE is intended to test the skills identified by the MacCrate Report as necessary for the practice of law by placing students in a practice and public-service-based setting for three months. Dean Glen suggested that the court system is the best location for this performance test because of its geographic accessibility, its inherent legitimacy, its need for additional assistance, and because “serving and improving the legal system through work in the courts’ justice initiatives is surely a means to promote and embody the MacCrate values.”¹⁹⁴ She also suggested that while engaged in public-service-

to Study the Bar Examination and Other Means of Measuring Lawyer Competence, Feb. 2012 at 14-5 [hereinafter NYSBA Report].

¹⁹¹ Those concerns included the following: (1) the difficulty of adding this proposal to a psychometrically validated exam; (2) the possibility students might not understand the benefit of clinical courses so the extra points may not help the students who most needed it; (3) the idea that instead of treating this as a bar exam issue, a clinic requirement could simply be made a licensing pre-requisite and required of all students. *Id.* at 15. California has proposed a competency training proposal that predicates a law license on bar applicants having completed, among other things, 15 units of practice-based, experiential course work or an apprenticeship equivalent during law school. Amy Yarborough, *Board Asks for Next Steps in Competency Training Plan*, CAL. ST. B.J. (Nov. 2013), <http://www.calbarjournal.com/November2013/TopHeadlines/TH4>. The proposal also requires “50 hours of legal services devoted to pro bono or modest means clients prior to admission or in the first two years of practice and 10 additional MCLE hours focused on law practice competency training.” *Id.*

¹⁹² NYBSA Report, *supra* note 190, at 16.

¹⁹³ Comms. on Legal Educ. & Admission to the Bar of the Ass’n of the Bar of the City of N.Y. & the N.Y. State Bar Ass’n, Public Service Alternative Bar Examination (June 14, 2002), http://www.nysba.org/Content/NavigationMenu/Attorney_Resources/NYSBA_Reports/JointRPT614.pdf; *see also* Glen, *Thinking out of the Bar Exam Box*, *supra* note 94; Glen, *Rethinking Admission*, *supra* note 1.

¹⁹⁴ Glen, *Rethinking Admission*, *supra* note 1, at 1723.

based lawyering, students would be tested not only on legal analysis, problem-solving, and written communication, but also on “oral and other forms of written communication, counseling, fact-gathering, familiarity with litigation and alternative dispute resolution, and time management. All of these skills would, of course, be utilized in the context of several bodies of substantive and procedural law, depending on the particular court.”¹⁹⁵ Students would be directly engaged in lawyering tasks and their lawyering skills would be supervised and assessed by trained evaluators in a realistic setting in which they would demonstrate what they can do, not just what they know. Crediting well-supervised practice experiences and offering alternative bar exams based on experiential work are ideas bar examiners should explore.

As with the New Hampshire program, these alternatives demand more human and other resources to accomplish than does the administration and scoring of a bar exam that measures only one form of intelligence and a subset of the lawyering skills necessary for practice. We do not propose that such alternatives necessarily replace the existing exam in its entirety, or that such proposals should be adopted in a comprehensive form in all jurisdictions. But neither should such proposals be abruptly dismissed or considered to be anti-intellectual and result-oriented efforts designed to promote diversity above competence. Rather, the consideration and development of such alternatives is a timely response to both the narrowness of our current testing regime and the troubling disparate impact it produces.

V. CRITIQUING THE LSAT

The LSAT, like the bar exam, has been a critical evaluative tool on the path into the legal profession.¹⁹⁶ As he does with respect to the bar

¹⁹⁵ Glen, *Thinking out of the Bar Exam Box*, *supra* note 94, at 426.

¹⁹⁶ Standard 503 requires use of “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 educational program.” Interpretation 503-1 requires that a law school using an admission test other than the LSAT “shall establish 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 educational program.” ABA SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, ABA STANDARDS AND RULES OF PROCEDURE FOR APPROVAL OF LAW SCHOOLS: 2014-2015 at 33 (2014) *available at* http://www.americanbar.org/groups/legal_education/resources/standards.html [hereinafter, ABA STANDARDS].

exam, Professor Subotnik suggests that questioning the use of the LSAT in the law school admissions process—as SALT has done¹⁹⁷—equates to valuing diversity over merit.¹⁹⁸ That view is based on a “general presumption that rank-ordering by test score aligns with rank-order admissions merit” so that “universities’ [or law schools’] reliance on non-test score, non-grade admissions criteria is assumed by many to be a deviation from a true academic merit-based standard.”¹⁹⁹ Because the LSAT’s presumed power to predict academic success underlies reliance upon the test as a “merit-based” admissions standard,²⁰⁰ we examine whether the LSAT is, in fact, the best predictor of academic success available. And because the purpose of law school is to prepare students to be lawyers, we consider how well the LSAT predicts success as a lawyer.

A. What’s Wrong with the LSAT?

As with the bar exam, the primary problem with the LSAT is the way it is used in decision-making, in this instance for law school admissions. The LSAT, like the SAT, is modeled on conventional tests

¹⁹⁷ Professor Subotnik notes that SALT has supported eliminating the LSAT as a required tool in law school admissions. Subotnik, *supra* note 2, at 379. In fact, SALT has primarily argued that the LSAT is misused in admissions. See Haddon & Post, *supra* note 1, at 104 (reprinting SALT statement). SALT also included an “if all else fails” argument:

If law schools continue to compete for distinction through popular magazine rankings, where high LSAT scores determine success; if there remains an unwillingness to challenge the perception that standardized tests measure innate intelligence; if those who administer admission programs continue to rely on the LSAT even when there is no correlation between test scores and either the performance of their students or the professional contributions of their graduates; if budgetary constraints are such that a careful, ‘whole file’ review system is regarded as prohibitively expensive and time-consuming, then it may be in the best interests of legal education to entirely abandon the Law School Admission Test.

Id. A decade later, with law schools relying on LSAT scores more than ever, perhaps it is almost time to follow that path, though we would prefer a more balanced and nuanced admissions process that considers the LSAT along with other factors and the development and reliance on entrance exams that test a broader array of relevant intelligence attributes. See *infra* Parts V.B, V. C.

¹⁹⁸ Subotnik, *supra* note, 2 at 385.

¹⁹⁹ West-Faulcon, *supra* note 8, at 1245.

²⁰⁰ *Id.* at 1281–82.

that measure general cognitive ability.²⁰¹ The test-taking skills and cognitive abilities tested on the LSAT are the same ones that are particularly rewarded in traditional first year law school courses, which typically assess a narrow range of analytical skills using multiple choice and essay or short answer questions administered under time pressure.²⁰² It is therefore not surprising that the LSAT is predictive of first year grades for some—but by no means all—students in doctrinal courses.²⁰³ But a single measure of cognitive ability such as the LSAT fails to fully predict academic performance,²⁰⁴ especially when that performance is assessed in ways that more closely replicate law practice.²⁰⁵ Moreover, the LSAT is not

²⁰¹ *Id.* at 1240; S. Newsome et al., *Assessing the Predictive Value of Emotional Intelligence*, 29 PERS. & INDIV. DIFFERENCES, 1005, 1008 (2000).

²⁰² *See generally* Shultz & Zedeck, *supra* note 1, at 623.

²⁰³ Abiel Wong, “Boalt-ing” Opportunity? *Deconstructing Elite Norms in Law School Admissions*, 6 GEO. J. ON POVERTY L. & POL’Y 199, 227 (1999) (noting that the LSAT’s correlation co-efficient with first-year grades ranges from .01 to .62 depending upon the law school); *see* Henderson, *supra* note 95, at 977; Haddon & Post, *supra* note 1, at 54–55.

“The [LSAT] test score, a product of one three-hour test, has a statistically significant correlation to first-year grades and is offered as a reliable predictor of whether an applicant will succeed in the first year of law school. But even this limited claim is contested, and the LSAC itself states that any predictive validity must be assessed on an individual school basis.”

Id.

²⁰⁴ Newsome, et al., *supra* note 201, at 1008 (noting that while cognitive tests predict 25% of the variance in academic performance, they leave 75% of that variance unexplained); *see also infra* sections V.B, VI.A (discussing the tests’ failure to accurately predict performance for a significant number of students and employees).

²⁰⁵ Curcio, Jones & Washington, *Does Practice Make Perfect? An Empirical Examination of the Impact of Practice Essays on Essay Exam Performance*, 35 FL. ST. L. REV. 271, 293 (2008) (finding that although LSAT scores correlated with first year grades in all doctrinal courses, there was no correlation with grades in the first year legal research and writing course); Henderson, *supra* note 95, at 976 (finding LSAT score correlated best with in-class timed exams, and had virtually no correlation to students’ grades on an appellate brief or oral advocacy assessment); Rolando Diaz et al., *Cognition, Anxiety, and Prediction of Performance in 1st Year Law Students*, 90 J. EDUC. PSYCHOL. 420, 423 (2001) (finding no correlation between LSAT scores and oral argument grades); *see also* Leah Christensen, *Enhancing Law School Success*, 33 L. & PSYCHOL.

designed to predict success in law practice,²⁰⁶ making it of limited value if law schools seek to admit applicants who will be successful lawyers, not just successful students. As Supreme Court Justice Samuel Alito has recognized, “[I]aw schools put too much emphasis on this one multiple choice test. What in life is a multiple choice test?”²⁰⁷

B. Learning from SAT-Optional Schools

Recognizing the limitations of standardized tests as predictors of college success, an increasing number of colleges and universities allow applicants the option of having their admissions applications evaluated without taking into account either SAT or ACT scores, a practice that began thirty years ago and has expanded considerably in the last decade. A recent study of 33 colleges and universities using this “test optional” policy found little difference in graduation rates or cumulative undergraduate GPAs between students who had submitted SAT or ACT scores [submitters] and those who were admitted into college without consideration of their SAT or ACT scores [non-submitters].²⁰⁸ While for both submitters and non-submitters, high school GPA was a consistent and reliable predictor of college cumulative GPA,²⁰⁹ for those not submitting SAT scores, the SAT scores were particularly unlikely to closely predict their college performance, even though the non-submitters’ SAT test scores were significantly lower than the scores of the submitters.²¹⁰

REV. 57, 74 (2009) (reporting results of a study in which grades in a lawyering skill class were better predictors of class rank than LSAT scores).

²⁰⁶ Shultz & Zedeck, *supra* note 1, at 621; Lempert et al., *From the Trenches and Towers*, 25 L. & SOC. INQUIRY 395, 401–02 (2000) (finding that LSAT scores and undergraduate grades had “virtually no value as predictors of post-law school accomplishments and success”).

²⁰⁷ Matthew Walther, *Sam Alito: A Civil Man*, THE AMERICAN SPECTATOR, May 2014, <http://spectator.org/articles/58731/sam-alito-civil-man>.

²⁰⁸ WILLIAM C HISS & VALERIE W. FRANKS, DEFINING PROMISE: OPTIONAL STANDARDIZED TESTING POLICIES IN AMERICAN COLLEGE AND UNIVERSITY ADMISSIONS at 3 (2014), <http://www.nacacnet.org/research/research-data/nacac-research/Documents/DefiningPromise.pdf>.

²⁰⁹ *Id.* at 9.

²¹⁰ *Id.* at 10. The scores of non-submitters were available in many instances because the institution either collected the scores of non-submitters for research purposes after admission or collected the scores from all applicants but kept them isolated from the rest of the admissions file for those to be evaluated without

This large study mirrors the findings of a review of twenty-five years of optional testing at Bates College, a highly-ranked and academically demanding institution that was an early adopter of test-optional admissions. Non-submitters earned the same grades and graduated at virtually the same rates as submitters.²¹¹ Both white and non-white students took substantial advantage of the policy, and the results for both groups were the same—no difference in grades and graduation rates between submitters and non-submitters.²¹² There was, however, a difference in the rates at which submitting and non-submitting students successfully pursued graduate degrees—especially where the career track required a high stakes standardized test for admission, (e.g. MBAs, PhDs, MDs and JD’s).²¹³ Given the nature of the Bates College study, there is no data to explain that gap (e.g., did non-submitting Bates graduates avoid such fields because of the entrance testing, or did they fail to apply or were rejected because of low test scores?), but the gap led the study’s authors (and the authors of this Article) to ask whether graduate schools are admitting the best candidates or just the best test-takers?²¹⁴

These studies raise profound questions about whether standardized tests should play a significant role in admissions decisions at any level. Test-optional colleges and universities have opened their doors to many who might not otherwise have applied or been accepted. The 33-institution survey, involving 123,000 students, found that non-submitters were more likely to be women, first-generation-to-college, all categories of minority students, Pell Grant recipients, and students with learning differences.²¹⁵ If such students perform as well in college as those with higher standardized test scores, as these two studies demonstrate, it suggests that using or heavily weighing cognitive tests

consideration of the scores. Looking at available scores at private colleges and universities, submitters had SAT scores 149 points higher than non-submitters. At public universities, in cases where students were admitted based on high school GPAs regardless of test scores, the non-submitters had SAT scores 93 points lower than the submitters. *Id.* at 13.

²¹¹ WILLIAM C. HISS & KATE M. DORIA, *DEFINING PROMISE: TWENTY-FIVE YEARS OF OPTIONAL TESTING AT BATES COLLEGE 1984-2009* at 7 (2011), <http://www.npr.org/assets/news/2013/optionaltestingpaper19842009.pdf>.

²¹² *Id.* at 5.

²¹³ *Id.* at 8–9.

²¹⁴ *Id.* at 9.

²¹⁵ HISS & FRANK, *supra* note 208, at 14.

like the SAT, ACT, and LSAT may be unfairly and unreasonably undermining the applications of many students who deserve to be, and might perform well if they are, admitted. While we know that the LSAT, especially in combination with undergraduate GPA, is somewhat predictive of law school grades, most of the variation in law school performance is not explained by those numbers.²¹⁶ Other recent studies have demonstrated that undergraduate performance is substantially affected by student perceptions of whether or not they “belong,” academically, at the institutions they attend and that performance dramatically improves when such perceptions are directly addressed.²¹⁷ Reliance on LSAT scores in the admissions process both obscures the limited nature of their predictive power and reinforces the counterproductive—and self-fulfilling—message to applicants with lower test scores that they are less likely to succeed in their studies. Perhaps it is time to allow law schools to choose test-optional admissions for at least some of their applicants, to invite in law schools a change that has produced strong positive results for undergraduate institutions and their students, especially those from diverse backgrounds.

When we suggest schools be allowed to experiment with “test optional” admissions we are not saying that the LSAT is irrelevant to

²¹⁶ David A. Thomas, *Predicting Academic Performance from LSAT Scores and Undergraduate Grade Point Averages: A Comprehensive Study*, 35 ARIZ. ST. L. J. 1007 (2003) (discussing the predictive power of LSAT scores and UGPAs as they relate to first year and cumulative grades of Brigham Young University students); see also Lisa Anthony Stilwell, et al., *Predictive Validity of the LSAT A National Summary of the 2009 and 2010 Correlation Studies*, LSAT TECHNICAL REPORT SERIES at 18 (2011), <http://www.lsat.org/docs/default-source/research-%28lsac-resources%29/tr-11-02.pdf?sfvrsn=2> [hereinafter LSAC study] (discussing findings of correlations between LSAT scores and first year law school GPAs [FYA] at 170 schools and finding significant variability in correlations between schools with an average correlation of LSAT scores and UGPA with FYA at .47; an average correlation of LSAT scores alone and FYA at .35 and .36; and a correlation of UGPA alone and FYA at .29 and .28)[hereinafter LSAC study]; Linda Wightman, *Beyond FYA: Analysis of the Utility of LSAT Scores and UGPA for Predicting Academic Success in Law School*, LAW SCHOOL ADMISSION COUNCIL RESEARCH REPORT 99-05, at 2 (2000) (finding a correlation between LSAT scores, UGPAs and cumulative law school grades).

²¹⁷ See Paul Tough, *Who Gets to Graduate?*, NEW YORK TIMES MAGAZINE (May 15, 2014), http://www.nytimes.com/2014/05/18/magazine/who-gets-to-graduate.html?_r=0.

or should not be used in admissions decisions for any students. Rather, as with the bar exam, the argument is that the LSAT has been misused because its submission by applicants is mandated by law school accreditation standards²¹⁸ and is relied on too heavily when it is used.²¹⁹ The Law School Admission Council (LSAC), which drafts and administers the LSAT, warns that LSAT scores should only be used as “one of several criteria for evaluation” of applicants²²⁰ and “should not be given undue weight solely because it is convenient.”²²¹ In an article based on research conducted while she was Vice President for Testing, Operations, and Research at the Law School Admission Council, Linda Wightman noted that “calling on [the LSAT] to do more than it was intended to do damages its validity. . . . [A] test that does a very good job of measuring a narrow, albeit important, range of acquired academic skills cannot serve as a sole determinant in the allocation of limited educational opportunity.”²²² It appears that many law schools ignore such warnings and instead overly rely upon LSAT scores, both for the sake of efficiency in reaching admissions decisions and because the *U.S. News and World Report Rankings* depend so heavily on the median LSAT scores of admitted students.²²³ Indeed, to

²¹⁸ Under Standard 503 of the ABA Standards and Rules of Procedure for Approval of Law Schools, schools must require each applicant to take “a valid and reliable admission test to assist the school and the applicant in assessing the applicant’s capability of successfully completing the school’s educational program.” ABA STANDARDS, *supra* note 196, at 33. If a school wants to use a test other than the LSAT for this purpose, it must establish that the test is a valid and reliable test to assist the school in this fashion.

²¹⁹ Haddon & Post, *supra* note 1, at 56–67.

²²⁰ Law School Admissions Council, *LSAC Statement of Good Admission and Financial Aid Practices- JD Program*, (May 2014), <http://www.lsac.org/docs/default-source/publications-%28lsac-resources%29/statementofgoodadm.pdf>.

²²¹ *Id.*

²²² Linda F. Wightman, *The Threat to Diversity in Legal Education: An Empirical Analysis of the Consequences of Abandoning Race as a Factor in Law School Admission Decisions*, 72 N.Y.U. L. REV. 1, 29–31 (1997).

²²³ Paula Lustbader, *Painting by the Numbers: The Art of Providing Inclusive Law School Admissions to Ensure Full Representation in the Profession*, 40 CAP. U. L. REV. 71, 114–117 (2012). LSAT scores account for 12.5 percent of the overall score used to rank law schools. Wendy Espland & Michael Sauder, *Rankings and Diversity*, 18 S. CAL. L. REV. & SOC. JUSTICE 587, 593 (2009). LSAT scores are “one of the very few [ranking] variables over which the law school has some input or control [and this] puts tremendous pressure on law schools to improve their median LSAT score to improve their relative rank.”

boost their rankings, schools have directed a significant amount of scholarship funding to induce students with higher LSAT scores to enroll.²²⁴ “Over-reliance on the LSAT thus has had the pernicious effect of reversing the long-established policy of offering scholarships to low-income students. This, in turn, contributes to the declining enrollment of low income and minority students.”²²⁵

Alex M. Johnson, Jr., *The Destruction of the Holistic Approach to Admissions*, 81 IND. L. J. 309, 311–12 (2006).

²²⁴ David Yellen, *The Impact of Rankings and Rules on Legal Education Reform*, 45 CONN. L. REV. 1389, 1395 (2013) (“The use of merit scholarships to attract students with high LSAT scores and UGPAs has exploded in the USNWR era, leaving far less money available for need-based aid.”); BRIAN Z. TAMANAHA, *FAILING LAW SCHOOLS* (Univ. of Chicago Press 2012) (observing that in a quest for US News rankings, law schools have shifted scholarship money from helping those who have significant financial needs to awarding the money to those with high LSAT scores); William D. Henderson & Andrew P Morris, *Student Quality As Measured by LSAT Scores*, 81 IND. L. J. 163, n. 5 (2006) (citing to comments by deans who note that they direct significant scholarship money to students with high LSAT scores). The practice of relying on standardized test scores as a proxy for merit in disbursing scholarship money is questionable given the recent studies that indicate students who do not submit test scores for undergraduate admissions perform as well as those with much higher test scores. See HISS & FRANK, *supra* note 208, at 15 (finding that only 5,046 out of 27,000 students who received merit awards were non-submitters with below-average-testing and yet that cohort of non-submitters earned slightly higher cumulative GPAs and graduated at rates 6% higher than submitters).

²²⁵ Kevin R. Johnson, *The Importance of Diversity of Students and Faculties in Law Schools, One Dean’s Perspective*, 96 IOWA L. REV. 1549, 1576–77 (2011) (“at a minimum, the U.S. News rankings methodology requires law school administrators to carefully weigh the ranking implications of any measures – such as less reliance on LSAT scores in admissions decisions designed to increase diversity among the student body.”); see also Law School Admissions Council, *LSAT As Predictors of Law School Performance* (2014), <http://www.lsac.org/docs/default-source/jd-docs/lSAT-score-predictors-of-performance.pdf>.

The correlation between LSAT scores and first-year law school grades varies from one law school to another (as does the correlation between GPA and first-year law school grades). During 2010, validity studies were conducted for 189 law schools. Correlations between LSAT scores and first-year law school grades ranged from .12 to .56 (median is .36).

Id.

C. Broadening the Admissions Process

While it once seemed to be generally believed that success in school could be at least somewhat predicted through measuring general intelligence,²²⁶ in recent years researchers have determined that successful performance depends on multiple kinds of intelligence. Instead of focusing solely on what now may be called “analytic intelligence,”²²⁷ they expand their vision to include other qualities such as “practical intelligence”²²⁸ and “creative intelligence.”²²⁹ Tests have been developed to assess these multiple aspects of intelligence, and the tests have been shown experimentally to better predict scholastic performance in college than tests focused on general intelligence.²³⁰ The tests of one researcher, Robert Sternberg, “were shown to have twice the practical predictive power of the SAT alone.”²³¹

Analytic, creative, and practical intelligence all seem likely to be useful attributes for learning and practicing law. However, like its SAT counterpart, the LSAT tests primarily analytical intelligence²³² and has

²²⁶ This attribute is sometimes called “g,” and is represented in exams such as the SAT and tests popularly known as IQ tests. See West-Faulcon, *supra* note 8, at 1256–64 (discussing the origin of the “g” factor tests and standardized tests’ influence on the development of IQ tests as well as the SAT). Although providing some mathematical basis for predicting success, SAT scores explain only about 13% of the variance in first-year college grades, leaving 87% of the variation unexplained by the test scores. See *id.* at 1267.

²²⁷ Sternberg, *Rainbow Project*, *supra* note 1, at 324 (“[the ability to] analyze, evaluate, judge, or compare and contrast”).

²²⁸ *Id.* at 325 (“[the ability to] implement, apply, or put into practice ideas in real-world context”).

²²⁹ *Id.* (“[the ability to] “create, invent, discover, support or hypothesize”).

²³⁰ See Karen Van Der Zee et al., *The Relationship of Emotional Intelligence with Academic Intelligence and the Big Five*, 16 EDUC. J. PERS. 103, 104 (2002) (noting the various measures that have been developed to assess social intelligence, practical intelligence and emotional intelligence); Sternberg, *Rainbow Project*, *supra* note 1, at 324–35.

²³¹ West-Falcon, *supra* note 8, at 1279.

²³² “The LSAT is a ‘paper-and-pencil’ test that basically measures analytic and logical reasoning, along with reading.” Shultz & Zedeck, *supra* note 1, at 622 (citing to the Law School Admissions Council 1999); see also Wendy M. Williams, *Consequences of How We Define and Assess Intelligence*, 2 PSYCHOL. PUB. POL’Y & L. 506, 509 (1996) (“Tests such as the Scholastic Assessment Test (SAT) and Preliminary Scholastic Assessment Test (PSAT), the GRE, the Law School Admission Test (LSAT), and the Graduate Management Admission

limited ability to predict who will succeed in law school²³³ or in law practice. With a goal of developing a better method of identifying which law school applicants have the qualities beyond cognitive ability needed to become responsible, effective and competent lawyers, Professors Marjorie Shultz and Sheldon Zedeck built upon the work done by Professor Sternberg and others to identify 26 lawyering “effectiveness” factors,²³⁴ to confirm the usefulness of the factors to explain lawyers’ on-the-job performance,²³⁵ and to find and develop assessment instruments that can test for non-cognitive factors such as interpersonal and communication skills, practical judgment, and creativity.²³⁶ Just as Professor Sternberg developed tests that could supplement the SAT to better predict college performance, Professors Shultz and Zedeck’s goal was to develop a law school admissions assessment instrument that could supplement the LSAT to better predict professional performance. The assessment measures they identified correlate significantly with the kinds of skills and qualities effective lawyers need.²³⁷

The primary goal for researchers like Professors Sternberg, Shultz, and Zedeck has been to develop better assessments so that admissions decisions based on merit can be made more accurately and therefore more fairly. They do not suggest eliminating the LSAT, but they argue that it should be supplemented with tests that measure a wider range of the skills and abilities effective lawyers need²³⁸ and better predict who

Test (GMAT), for example, all measure verbal and mathematical knowledge and reasoning.”).

²³³ See Law School Admission Council, *supra* note 225; see also Sternberg, *Rainbow Project*, *supra* note 1, at 324 (noting that traditional tests primarily measure analytical skills, but that success is attained through a balance of analytical, practical, and creative intelligence).

²³⁴ Shultz & Zedeck, *supra* note 1, at 630.

²³⁵ *Id.* at 637–638.

²³⁶ *Id.* at 625.

²³⁷ *Id.* at 642.

²³⁸ *Id.* at 630. Similar efforts are being or have been considered in graduate, business, and medical school admissions. See, e.g., Association of American Medical Colleges, *MR5: 5th Comprehensive Review of the Medical College Admission Test (MCAT)*, AAMC (2011), <https://www.aamc.org/initiatives/mr5>; Association of American Medical Colleges, *Innovation Lab to Explore Measures of Personal Characteristics and Skills*, AAMC (2009), https://www.aamc.org/initiatives/mr5/about_mr5/64636/innovation_lab.html; Alison Damast, *The GMAT Gets a Makeover*, BLOOMBERG BUSINESSWEEK

will succeed. The broader assessments proposed for both college and law school admissions have the added benefit of producing fewer performance gaps by race and therefore diminish the disparate impact of relying exclusively on the analytic measurements. The Shultz-Zedeck tests, for instance, “showed few racial or gender subgroup differences, creating the potential to reduce adverse impact through the use of new tests.”²³⁹

Professor Subotnik dismisses the Shultz-Zedeck research on the grounds that the lawyering effectiveness factors do not encompass qualities such as “learned in the law.”²⁴⁰ But knowing the law—if that is what Professor Subotnik means when he uses that term—is part of the many Shultz-Zedeck effectiveness factors such as skills in legal analysis and reasoning, problem-solving, and researching the law.²⁴¹ The work of Professors Sternberg, Shultz, and Zedeck represent developing, not definitive, responses to the limitations of current admissions testing and its disparate impact. They have shown that testing alternatives exist, which, when added to existing cognitive tests, offer the possibility of predicting ability to succeed as a lawyer better than the primarily or solely cognitive tests now in use and they do so without the disparate impact produced by current testing. It is time to rethink the assumption that earning a higher score on a “particular mental test” indicates that the test-taker is more qualified

(June 4, 2010), http://www.businessweek.com/bschools/content/jun2010/bs20100624_048037.htm; Lucia Graves, *More Changes Ahead for the GRE Test*, U.S. NEWS & WORLD REPORT (May 27, 2008, 5:35), <http://www.usnews.com/education/blogs/on-education/2008/05/27/more-change-s-ahead-for-the-gre-test>.

²³⁹ *Id.* at 654.

²⁴⁰ Subotnik, *supra* note 2, at 383. Professor Subotnik also dismissed Schultz and Zedeck’s work on the ground that their tests “may be developed through instruction.” *Id.* at 383. Yet, even if taken as true, this does not distinguish it from existing tests such as the SAT, LSAT and Bar Exam, which we know are coachable and the basis of a lucrative test-preparation industry. Those who can afford to pay for private coaches can achieve significant test score improvements. See JOHN P. HAUSKNECHT ET AL., RETESTING IN SELECTION: A META-ANALYSIS OF PRACTICE EFFECTS FOR TESTS OF COGNITIVE ABILITY (2006), <http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1012&context=articles>.

²⁴¹ Shultz & Zedeck, *supra* note 1, at 630 (listing the factors identified as important to lawyer effectiveness).

than someone earning a lower score but perhaps possessing untested qualifications to succeed as an effective, practicing lawyer.²⁴²

VI. TEST-TAKING AND ECONOMICS

Professor Subotnik argues that when we ignore tests of cognitive abilities, we enter economically perilous territory,²⁴³ pointing to studies correlating job performance to general intelligence (“g”) test scores, as well as findings from a study undertaken by the Organization of Economic Co-operation and Development [OECD].²⁴⁴ In this section, we address those arguments.

A. Cognitive Tests and Job Performance

Professor Subotnik argues that in a quest for a diverse workforce, America has turned into a country in which diversity trumps job qualifications.²⁴⁵ He cites Professor Amy Wax,²⁴⁶ who reviews studies about the predictive force of cognitive tests as they relate to job performance,²⁴⁷ and argues that employers are hiring unqualified workers due to fear of Title VII litigation.²⁴⁸ Using Professor Wax’s work to buttress his arguments, Professor Subotnik suggests employers are forced to choose unqualified diverse workers, which negatively impacts the economy.²⁴⁹

This argument rests on presumptions that cognitive tests alone best predict job performance and that diversity considerations require ignoring rank order on cognitive tests when making hiring decisions,

²⁴² See West-Faulcon, *supra* note 8, at 1295 (arguing that fairness in selective admissions processes requires one to consider that the scientific data does not support the assumption that those with higher test scores on cognitive tests are necessarily the most qualified applicants).

²⁴³ Subotnik, *supra* note 2, at 344–45, 392–94.

²⁴⁴ *Id.* at 393.

²⁴⁵ *Id.* at 361.

²⁴⁶ Amy L. Wax, *Disparate Impact Realism*, 53 WM. & MARY L. REV. 621 (2011).

²⁴⁷ Subotnik, *supra* note 2, at 356.

²⁴⁸ *Id.* at 359 (citing Wax *supra* note 246, at 694–95).

²⁴⁹ See *id.* at 361 (noting “[a]t the macro level in the United States there is, unavoidably in Wax’s words, a ‘validity-diversity tradeoff’”); see also *id.* at 389–94; see generally *id.* at 359–364 (using Professor Wax’s work to suggest that jobs are going to unqualified workers because employers seek to avoid litigation and create a diverse workforce).

resulting in the hiring of less qualified workers. That analysis is flawed in several ways.

First, while cognitive tests have some value in predicting job performance—as we have seen with respect to the firefighting test in New Haven, the bar exam, and the LSAT²⁵⁰—they are far from perfect predictors. As Professor Subotnik himself acknowledges, Professor Wax’s estimate that the tests predict job performance for approximately fifty percent of the candidates is likely an over-estimation.²⁵¹ Many researchers suggest that cognitive tests predict performance for about twenty to twenty-five percent of the workforce,²⁵² while others put the tests’ predictive ability even lower.²⁵³ If the tests predict performance for twenty to twenty-five percent of workers, then for close to four out of five job candidates, cognitive test scores do not accurately predict ability to succeed in a given job.

Second, as we have also seen with respect to the legal profession, cognitive abilities are not all that it takes to succeed in the workplace. Practical problems encountered on the job often require the ability to recognize problems that are ill-defined, require information-seeking, may have multiple solutions and multiple paths to a solution, may require reliance on information learned in every-day experience, and potentially require motivation and personal involvement, a different set of skills than those involved in solving academic problems.²⁵⁴ Abilities such as interpersonal communication skills, practical judgment, and creativity play a role in successful job performance.²⁵⁵

²⁵⁰ Sternberg & Hedlund, *supra* note 1, at 143.

²⁵¹ Subotnik, *supra* note 2, at 360.

²⁵² Sternberg & Hedlund, *supra* note 1, at 144.

²⁵³ See Christopher Jencks & Meredith Phillips, *Introduction to THE BLACK-WHITE TEST SCORE GAP* 1, 15 (Christopher Jencks & Meredith Phillips eds., Brookings Inst. Press 1998) (noting that “test scores explain only 10 to 20 percent of the variation in job performance”); Wendy M. Williams, *Consequences of How We Define and Assess Intelligence*, 2 *PSYCHOL. PUB. POL’Y & L.* 506, 511 (1996) (“it is clear that between 75% and 96% of the variance in real-world criteria such as job performance cannot be accounted for by individual differences in intelligence test scores”).

²⁵⁴ Robert J. Sternberg & Richard K. Wagner, *The g-centric View of Intelligence and Job Performance is Wrong*, 2 *CURRENT DIRECTIONS IN PSYCHOL. SCI.* 1, 2 (1993).

²⁵⁵ Shultz & Zedeck, *supra* note 1 at 625; see generally Flip Lievens & David Chan, *Practical Intelligence, Emotional Intelligence and Social Intelligence*, in

Professors Barrick and Mount found that conscientiousness, extraversion, emotional stability, agreeableness, and openness to experience relate to job performance, with the strength of that relationship depending upon the occupation and the job task.²⁵⁶ Other studies have found correlations between job performance and emotional²⁵⁷ and practical intelligence.²⁵⁸ Simply put, the best workers may not always be those who performed best on cognitive tests.²⁵⁹ If those tests are used to decide who gets hired, companies may fail to hire someone who is well-qualified to perform the job and adds an important perspective to the work team.

HANDBOOK OF EMPLOYEE SELECTION 339 (James L. Farr & Nancy T. Tippins eds., 2010) (discussing the various conceptualizations of intelligence, the instruments used to measure these intelligences, and the correlation between these various conceptualizations of intelligence and job performance prediction).

²⁵⁶ Murray R. Barrick & Michael K. Mount, *The Big Five Personality Dimensions and Job Performance: A Meta-Analysis*, 44 PERSONNEL PSYCHOL. 1, 1 (1991); see also Nathan R. Kuncel et. al., *Individual Differences as Predictors of Work, Educational and Broad Life Outcomes*, 49 PERSONALITY & INDIVIDUAL DIFFERENCES 331, 334–35 (2010) (discussing various studies that indicate personality factors can be assessed and the assessments indicate a strong correlation between various personality factors and job performance).

²⁵⁷ See generally Lievens & Chan, *supra* note 255. Lievens and Chan describe two types of emotional intelligence. One type is assessed via performance based tests and deals with the ability to accurately perceive others' emotions from behavioral and other non-verbal cues, to use emotions to assist in thinking and problem solving, to analyze emotions and think about how they affect outcomes, and to manage one's emotions. *Id.* at 340–41. The other type of emotional intelligence is measured via self-reporting instruments that assess the ability to “recognize, understand, and use emotional information about oneself or others that leads to or causes effective or superior performance.” *Id.* at 341. See also Shaun Newsome et. al., *Assessing the Predictive Value of Emotional Intelligence*, 29 PERSONALITY & INDIVIDUAL DIFFERENCES 1005 (2000).

²⁵⁸ Professor Sternberg coined the term “practical intelligence” in order to describe how people apply their abilities to real world problems they confront on the job or at home. See Sternberg, *Rainbow Project*, *supra* note 1, at 325.

²⁵⁹ See Sternberg & Hedlund, *supra* note 1 at 144 (reviewing literature that discusses limitations and controversies surrounding sole reliance on “general intelligence” to predict job performance); see also Karen Van der Zee et al., *The Relationship of Emotional Intelligence with Academic Intelligence and the Big Five*, 16 EUR. J. PERS. 103, 103 (2002) (noting that sometimes those who do very well in school are unsuccessful at work, despite having strong intellectual abilities).

B. The Role of Unconscious Bias in Workplace Performance Evaluations

The arguments about the connection between cognitive test scores and job performance also ignore the role bias plays in workplace evaluations. Professor Subotnik contends that lower test scores correlate to lower job performance, that measures of job performance strongly correlate to “objective” measures such as work errors, and that racial bias plays only a marginal role in evaluation of workplace performance. He concludes that employers hire less effective workers when they hire diverse candidates with lower cognitive test scores.²⁶⁰ A recent study demonstrates that implicit racial bias,²⁶¹ combined with confirmation bias, may in fact play a much larger role in workplace evaluations than acknowledged by Professor Subotnik or the studies on which he relies.

Confirmation bias is the often unconscious tendency to seek or interpret evidence in ways that conform to one’s existing beliefs or expectations.²⁶² Confirmation bias explains how our unconscious perceptions affect our evaluations – e.g. we see weaker performance when we expect to see weaker performance and vice versa.²⁶³ If one holds unconsciously biased attitudes about a group of people, one’s seemingly objective assessment of people in that group may be tainted by confirmation bias. A recent study illustrates how confirmation bias in the legal workplace may result in lower evaluations for African-American associates than for Caucasian associates for exactly the same work product.²⁶⁴ Five partners from five different firms drafted a

²⁶⁰ Subotnik, *supra* note 2, at 363–64.

²⁶¹ L. Song Richardson, *Cognitive Bias, Police Character and the Fourth Amendment*, 44 ARIZ. ST. L. J. 267, 271 (2012) (“Implicit racial bias describes a psychological process in which a person’s non-conscious racial beliefs (stereotypes) and attitudes (prejudices) affect her or his behaviors, perceptions and judgments in ways that she or he are largely unaware of and typically, unable to control.”).

²⁶² *See generally*, Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. OF GEN. PSYCHOL 175 (1998) (explaining confirmation biases and reviewing evidence of, and explanations for, this bias).

²⁶³ *Id.*

²⁶⁴ NEXTIONS LLC, YELLOW PAPER SERIES: WRITTEN IN BLACK & WHITE EXPLORING CONFIRMATION BIAS IN RACIALIZED PERCEPTIONS OF WRITING SKILLS at 5 (2014), http://www.nextions.com/wp-content/files_mf/13972237592014040114WritteninBlackandWhiteYPS.pdf.

legal research memo, deliberately inserting grammatical, factual and analytical errors.²⁶⁵ The memo then was analyzed by 53 other law firm partners who agreed to participate in a “writing analysis study.”²⁶⁶ All were told the memo was drafted by a male third-year associate who matriculated from NYU Law School. Half were told the associate was Caucasian; the other half were told the associate was African American.²⁶⁷ The identical memo averaged statistically-significant lower ratings for the African American associate than for the Caucasian associate.²⁶⁸ The qualitative comments were also more positive for the Caucasian memo writer,²⁶⁹ and, on average, the reviewers found significantly more errors in the memo written by the “African American” than they found in the memo written by the “Caucasian” writer.²⁷⁰ The underlying results—that implicit and confirmation bias appear to play a role in workplace evaluations—appeared also in another study in which minority summer associates were consistently evaluated more negatively than their majority counterparts; the researchers found that blind evaluations of work product were generally more positive for minorities and women and less positive for majority men than when the identity of the individuals being evaluated was known.²⁷¹

If evaluators unconsciously find more errors under the influence of implicit and confirmation bias, this will affect their “objective” evaluation of the quality of the work produced. The conclusions that workplace performance confirms the validity of cognitive testing and

²⁶⁵ *Id.*

²⁶⁶ Originally 60 different partners were recruited: 23 women, 37 men, 21 self-identifying as being a racial/ethnic minority and 39 self-identifying as Caucasian. Of those 60, 53 completed the editing and rating of the memo. *Id.* at 2.

²⁶⁷ *Id.*

²⁶⁸ *Id.* at 3.

²⁶⁹ The comments about the Caucasian writer were: “generally good writer but needs to work . . .” ; “has potential”; “good analytical skills”. The African-American writer received comments such as: “needs lots of work”; “can’t believe he went to NYU”; “average at best.” *Id.* at 4.

²⁷⁰ For the “African American” writer, on average reviewers found 5.8/7.0 spelling/grammar errors; 4.9/6.0 technical writing errors, and 3.9/5.0 errors in facts. For the “Caucasian” writer, reviewers found 2.9/7.0 spelling/grammar errors; 4.1/6.0 technical writing errors and 3.2/5.0 errors in facts. *Id.* at 5.

²⁷¹ *Id.* at 5.

that discounting cognitive tests undermines economic productivity thus rest on suspect assumptions about the objectivity of workplace performance measures.

C. The Economic Value of Diversity

Finally, the argument that diversity concerns lead employers to make “wrong” decisions about hiring ignores the economic value to employers of diversity in the workplace. Many employers find that a diverse work force creates a competitive marketplace advantage because it helps companies better understand and serve a growingly diverse customer base.²⁷² As noted by Nancy Levit five years ago, “[n]umerous studies, both in this country and others, have demonstrated that ‘diversity is good for business.’ . . . The market arguments in favor of diversity are compelling.”²⁷³

The same is true for the legal system, as recognized by the ABA in a 2010 report developed as the result of an ABA Presidential Diversity Initiative. Articulating democracy, business, leadership, and demographic rationales for diversity, the report noted that

[w]ithout a diverse bench and bar, the rule of law is weakened as the people see and come to distrust their exclusion from the mechanisms of justice. . . . A diverse workforce within legal and judicial offices exhibits different perspectives, life experiences, linguistic and cultural skills, and knowledge about international markets, legal regimes, different geographies, and current events. . . . It makes good business sense to hire lawyers who reflect the diversity of citizens, clients, and customers from around the globe. Indeed, corporate clients increasingly require lawyer diversity and will take their business elsewhere if it is not provided. . . . As Justice O’Connor reminded us in her opinion in the Grutter case, this society draws its leaders from the ranks of the legal profession and that is one reason why diversity is a constitutionally protected principle and practice.²⁷⁴

In the legal profession, as elsewhere, economic, and other interests, are often best served by having a diverse workforce.²⁷⁵

²⁷² Nancy Levit, *Megacases, Diversity, and the Elusive Goal of Workplace Reform*, 49 B.C. L. REV. 367, 424–27 (2008).

²⁷³ *Id.* at 424–25.

²⁷⁴ AMERICAN BAR ASSOCIATION, *DIVERSITY IN THE LEGAL PROFESSION: THE NEXT STEPS, REPORT AND RECOMMENDATIONS* 9–10 (2010).

²⁷⁵ Cruz Reynoso & Cory Amron, *Diversity in Legal Education: A Broader View, A Deeper Commitment*, 52 J. LEGAL EDUC. 491, 505 (2002).

In sum, the connection between cognitive test scores and job performance is not nearly as strong as Professor Subotnik assumes. When cognitive tests are supplemented with assessments of other factors, such as practical or emotional intelligence, employers can better predict who will be a successful employee—as well as minimize the disparate impact of the racial gap in test scores.²⁷⁶ It is not an “all or nothing” question, as Professor Subotnik suggests. Rather, like the arguments about tests that determine who will be a lawyer, it is a question of how much weight should be given to tests that do not perfectly predict performance, especially in light of potentially viable alternatives that can supplement the tests and improve their predictive abilities.

D. OECD study

Professor Subotnik supports his argument that we ignore cognitive test scores at our economic peril in part by citing an OECD²⁷⁷ study, which suggests a relationship between cognitive skills of a nation’s student population and that nation’s Gross Domestic Product (GDP).²⁷⁸ The OECD report is part of the OECD’s Programme for International Student Assessment (PISA) which “represents a commitment by governments to monitor outcomes of education systems in terms of student achievement, within a common international framework.”²⁷⁹ PISA attempts to assess young people’s reading, math and science literacy skills, across national boundaries.²⁸⁰ The OECD report, to which Professor Subotnik cites, compared 15

²⁷⁶ Shultz & Zedeck, *supra* note 1, at 625; Sternberg & Hedlund, *supra* note 1, at 153; *see also* James Clevenger et. al, *Incremental Validity of Situational Judgment Tests*, 86 J. APPLIED PSYCHOL. 410 (2001) (finding situational judgment tests were a valuable additional predictor of job performance).

²⁷⁷ Subotnik, *supra* note 2, at 393; *see also* ORGANIZATION OF ECONOMIC CO-OPERATION AND DEVELOPMENT, THE HIGH COST OF LOW EDUCATIONAL PERFORMANCE: THE LONG-RUN IMPACT OF IMPROVING PISA SCORES (2010), *available at* <http://www.oecd.org/pisa/44417824.pdf> [hereinafter OECD REPORT] (assessing the test scores of fifteen year olds from participating countries).

²⁷⁸ OECD REPORT, *supra* note 277, at 16–17.

²⁷⁹ *Id.* at 3.

²⁸⁰ ORGANIZATION OF ECONOMIC CO-OPERATION AND DEVELOPMENT, MEASURING STUDENT KNOWLEDGE AND SKILLS: A NEW FRAMEWORK FOR ASSESSMENT 9 (1999), *available at* <http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33693997.pdf>.

year olds' test scores with their countries' GDP and found a correlation between cognitive math and science scores and GDP.²⁸¹ They then extrapolated that finding to suggest that improving cognitive test scores would result in higher GDP.²⁸² That claim is open to question.

The test methodology underlying the OECD Report has been criticized by numerous academics.²⁸³ Moreover, the claims that test scores relate to GDP and that raising test scores can lead to an increase in GDP are controversial. As the OECD researchers themselves note, even if cognitive skill levels, as measured by the PISA test, correlate to GDP it is hard to prove that higher cognitive skill levels produce higher GDP.²⁸⁴ Two studies found that when one controls for variables typically used when making international comparisons, the strong association between PISA cognitive test scores and GDP per capita disappeared.²⁸⁵ GDP likely depends upon multiple variables working together.²⁸⁶ Again, we do not argue that cognitive skills are unimportant. We, and others who criticize overreliance on standardized tests, are not “undermining the case for the highest educational standards.”²⁸⁷ Rather, we suggest that it is a mistake to over-emphasize the value of standardized cognitive tests in considering how to improve GDP, just as it is a mistake to look at only cognitive test results to determine who is likely to succeed in school or a job. We should instead recognize that a country's economic success is based upon many different factors, only some of which may be measured by particular tests.

²⁸¹ OECD REPORT, *supra* note 277, at 17.

²⁸² *Id.* at 6.

²⁸³ See, e.g., William Stewart, *Is Pisa Fundamentally Flawed*, TESCONNECT (Dec. 4, 2013), <http://www.tes.co.uk/article.aspx?storycode=6344672> (outlining various critiques of the test and noting that the test questions used vary between students and between countries, thus rendering comparisons meaningless).

²⁸⁴ OECD REPORT, *supra* note 277, at 17–20 (reviewing issues relating to causality).

²⁸⁵ See Shiu Sheng Chen & Ming-Ching Luoh, *Are Mathematics and Science Test Scores Good Indicators of Labor Force Quality?*, 96 SOC. INDICATOR RES., 96, 133, 138–143 (2010); Chong Ho Yu, Samuel DiGangi, & Angel Jannasch-Pennell et al., *A Time-Lag Analysis of the Relationships Among PISA Scores, Scientific Research Publication, and Economic Performance*, SOC. INDICATOR RES. 317, 329–30 (2012).

²⁸⁶ Yu, et al., *supra* note 285, at 329.

²⁸⁷ See Subotnik, *supra* note 2, at 394.

VII. TEST-TAKING AND RACIAL GAPS

Underlying the argument about the meaning of and appropriate response to racial gaps in test results is the acknowledgment and understanding of the gap itself. We have explained elsewhere that the test scores should be used only as part of a broader decision-making paradigm. But what does the existence of the gap suggest about the value of the tests themselves? We end this Article with a discussion of explanations offered for the test-score gaps between white and minority test-takers to reinforce our conclusion that decision-makers should not use the scores as a decisive factor for hiring, promotion, school admission, or licensing without exploring more fully the validity and scope of the tests, and considering alternative ways to measure merit.

The difference between results for white and minority—especially African-American—test-takers is seen as early as pre-school and continues through adulthood.²⁸⁸ The reasons for the gaps remain elusive, as does the remedy. Some assert definitively dismissed explanations based on belief in inherent genetic differences between different racial populations.²⁸⁹ Some suggest societal explanations, including subpar schools,²⁹⁰ inadequately prepared teachers,²⁹¹ insufficiently challenging curricula,²⁹² unequal treatment by and

²⁸⁸ Jencks & Phillips, *supra* note 253, at 1.

²⁸⁹ See generally RICHARD HERRNSTEIN & CHARLES MURRAY, *THE BELL CURVE: INTELLIGENCE AND CLASS STRUCTURE IN AMERICAN LIFE* (1994); see also Richard E. Nesbitt, *Race, Genetics and IQ*, in *THE BLACK-WHITE TEST SCORE GAP*, *supra* note 253, at 86–101.

²⁹⁰ Roy L. Brooks, *American Democracy and Higher Education for Black Americans: The Lingering Effects Theory*, 7 *J.L. SOC. CHALLENGES* 1, 65 (2005).

²⁹¹ See generally LINDA DARLING-HAMMOND, *TEACHER QUALITY AND STUDENT ACHIEVEMENT: A REVIEW OF STATE POLICY EVIDENCE* (1999), available at <http://www.politicalscience.uncc.edu/godwink/PPOL8687/WK11March%2029%20Teachers/Darling-Hammond%20Review%20essay%20on%20teacher%20quality%20and%20outcomes.pdf> (discussing how increasing investment in teacher quality leads to improved student performance and how economically disadvantaged and minority students have limited access to the most qualified teachers).

²⁹² Brooks, *supra* note 290, at 65 (noting that black students are less likely to be enrolled in AP and honors level courses than their white counterparts).

expectations from teachers,²⁹³ weak or missing educational support at home,²⁹⁴ and socioeconomic differences.²⁹⁵ Some argue that the problem is the failure of African American youth to work as hard as their white counterparts because due to a history of discrimination they “began to doubt their own intellectual ability, began to define academic success as white people’s prerogative and began to discourage their peers, perhaps unconsciously, from emulating white people in academic striving, i.e. from ‘acting white.’”²⁹⁶ Some suggest that test deficiencies cause some of the disparity—that is, that the tests themselves are racially biased.²⁹⁷

Another explanation is “stereotype threat,” a term coined by Claude Steele and Joshua Aronson to explain their findings that “the existence of a negative stereotype about a group to which one belongs means that in situations where it is potentially applicable, one risks

²⁹³ See generally Ronald F. Ferguson, *Teachers’ Perceptions and Expectations and the Black-White Test Score Gap*, in *THE BLACK-WHITE TEST SCORE GAP*, *supra* note 253, at 273–314 (discussing various studies about teachers’ perceptions and expectations and concluding that “teacher’s perceptions, expectations, and behaviors probably do help to sustain, and perhaps even to expand, the black-white test score gap”).

²⁹⁴ See generally Meredith Phillips et al, *Family Background, Parenting Practices, and the Black-White Test Score Gap*, in *THE BLACK-WHITE TEST SCORE GAP*, *supra* note 253, at 103–48 (suggesting that parenting practices play a role in test score differences).

²⁹⁵ Larry V. Hedges & Amy Nowell, *Black-White Test Score Convergence Since 1965*, in *THE BLACK-WHITE TEST SCORE GAP*, *supra* note 253, at 149–81 (reviewing the literature and statistics and concluding that historically, parental education and income have played a role in partially explaining test score gaps).

²⁹⁶ Signithia Fordham & John U. Ogbu, *Black Students’ School Success: Coping with the “Burden of ‘Acting White’”*, 18 *URB. REV.* 176, 177 (1986). This theory has been, at least implicitly, endorsed by Professor Subotnik. See Subotnik, *supra* note 2, at 388 (suggesting grades are a result of ambition, “stick-to-it-iveness, and attention to detail”); *id.* at 389 (“we do not know how to motivate our black youth to earn high grades”). The theory has been discounted by other social scientists. See Philip J. Cook & Jens Ludwig, *The Burden of “Acting White”: Do Black Adolescents Disparage Academic Achievement*, in *THE BLACK-WHITE TEST SCORE GAP*, *supra* note 253, at 375–400. Professors Cook and Ludwig’s conclusions based upon the National Educational Longitudinal Study have been challenged by Professor Ferguson—a challenge that the authors respond to in their chapter. *Id.* at 394–98.

²⁹⁷ Jencks & Phillips, *supra* note 253, at 12–15.

confirming that stereotype, both to oneself and to others.”²⁹⁸ Through a series of experiments they confirmed that “making African Americans more conscious of negative stereotypes about their intellectual ability as a group can depress their test performance relative to that of whites.”²⁹⁹ A statement to black Stanford students as seemingly innocuous as telling them they were about to take a diagnostic test of their academic abilities led the students to perform significantly worse than when the same test was described as a laboratory problem-solving task.³⁰⁰ They found that merely asking African American test-takers to report their race was enough to impair their performance, even when the test was not described as a measure of ability.³⁰¹ Steele and Aronson found that “[S]tereotype threat seems to exert its influence by reducing efficiency. Participants who experience stereotype threat spend more time doing fewer items less accurately.”³⁰² Their work suggests stereotype threat lowers the scores of even, or perhaps especially, high achieving black students on verbal tests akin to the SAT,³⁰³ and has implications that may help explain the test score gaps on tests such as the SAT, LSAT, GRE, and even the bar exam.³⁰⁴ The existence of stereotype threat has been supported by a wide range of empirical research on everything from performance on math tests by women and white men³⁰⁵ to academic performance by athletes.³⁰⁶

²⁹⁸ Claude M. Steele & Joshua Aronson, *Stereotype Threat and the Test Performance of Academically Successful African Americans*, in *THE BLACK-WHITE TEST SCORE GAP*, *supra* note 253, at 422.

²⁹⁹ *Id.* at 422.

³⁰⁰ *Id.* at 409.

³⁰¹ *Id.* at 421 (discussing the results of a difficult verbal test).

³⁰² *Id.* at 423.

³⁰³ *See id.* at 424 (discussing how the results from their tests relate to the SAT scores of the test subjects).

³⁰⁴ *See, e.g., id.* at 425 (“[W]ould the results hold for other kinds of tests? . . . Our theory predicts that one would find the same results. . .”).

³⁰⁵ Joshua Aronson et al., *When White Men Can't Do Math: Necessary and Sufficient Factors in Stereotype Threat*, 35 *J. EXPERIMENTAL SOC. PSYCHOL.* 29, 29 (1999) (finding that activating stereotypes produced lower performance by white males on math tests); *cf.* Catherine Good et al., *Problems in the Pipeline: Stereotype Threat and Women's Achievement in High-Level Math Courses*, 29 *J. APPLIED DEVELOPMENTAL PSYCHOL.* 17, 25 (2008) (finding that even amongst highly qualified women, stereotype threat affects test performance).

³⁰⁶ *See* Jeff Stone, *A Hidden Toxicity in the Term “Student-Athlete”: Stereotype Threat for Athletes in the College Classroom*, 2 *WAKE FOREST J. L. & POL’Y*

Whatever the explanation for the test gap on standardized tests, the important point is that the most likely explanation is not inherent inferiority of those who perform less well, but some combination of factors that depress performance on cognitive tests. If cognitive test performance were a perfect—or even a very good—predictor of performance, perhaps the response should be remedial work to improve performance on the test itself. But we know that cognitive tests are far from perfect predictors of the best performers, and certainly far from infallible in separating those who can succeed from those who can't. That is true for academic performance³⁰⁷ as well as for job performance.³⁰⁸ Indeed, Jencks and Phillips note that “test scores explain only 10 to 20 percent of the variation in job performance” and, of equal importance, that “blacks are far less disadvantaged on the non-cognitive determinants of job performance than on the cognitive ones.”³⁰⁹ In a statement echoing our own claims in this article, they write that

if racial fairness means that blacks and whites who could do a job equally well must have an equal chance of getting the job, a selection system that emphasizes test scores is almost always unfair to most blacks (and to everyone else with low test scores). . . . [Relying on cognitive test scores] forces blacks to pay for the fact that social scientists have unusually good measure of a trait on which blacks are unusually disadvantaged.³¹⁰

179 (2012) (discussing the role stereotype threat may play on academic performance of those designated “student-athletes”).

³⁰⁷ Brooks, *supra* note 290, at 63–64 (citing Christopher Jencks, *Racial Bias in Testing*, in *THE BLACK-WHITE TEST SCORE GAP*, *supra* note 253, at 65).

Many students, black or white, at academically demanding colleges outperform their SAT scores. For example, in one study of such schools, students admitted with an ‘unacceptable’ SAT score of 1,000 ‘earned grades averaging roughly B’, while students admitted with scores at the school-wide average of 1289 ‘earned college grades averaging roughly B+ Anyone who has graded undergraduates knows that this is not a large difference. Indeed, it is the smallest difference that most college grading systems bother to record.’

Id.

³⁰⁸ Jencks & Phillips, *supra* note 253, at 14–15.

³⁰⁹ *Id.* at 15.

³¹⁰ *Id.*

If we know of, or can develop, tests that account for non-cognitive factors that help lead to and explain success, fairness dictates that we expand our test-taking universe to encompass them—especially if those additional tests reduce the race gap in admissions, licensing, and employment decisions.

VIII. CONCLUSION

Given the existing gaps in cognitive test scores—including scores on the LSAT and bar exam—and their inadequacy in assessing merit, the critical question is whether we award admissions slots and jobs based on those test results, or instead look beyond cognitive tests to develop better methods of assessing who is likely to succeed. Professor Subotnik’s view is that the goal should be to raise African American test scores so that they can “beat white folks at their own game.”³¹¹ That is a far too limited solution, and one that is premised upon an unreasonable, unfair, and unwise deference to cognitive-based testing and a flawed belief that failing to rely on cognitive test scores in allocating key human capital resources, such as admissions slots and jobs, is unfair to white people and economically harmful to the United States. While addressing systemic issues that lead to lower scores is important, we must limit our reliance on cognitive test scores that imperfectly predict success and fail to account for other identifiable factors that are important to success. The unfair and troubling disparate impact of cognitive tests is a warning signal. Our response should not be to double down on those tests, but instead to develop better measures of merit.

Professor Subotnik argues that questioning the value placed upon cognitive-based tests in the face of their disparate impact is anti-intellectual and represents a “race comes first” approach.³¹² As we have discussed throughout this Article, that argument adopts an ill-informed view of merit. Cognitive-based tests are far from perfect predictors of success in school, in a job, or in law practice. They fail to account for many abilities and qualities that lead to successful performance. While we do not suggest cognitive abilities are unimportant, placing undue weight upon cognitive-based tests ill-serves society. Although it is perhaps easy and efficient, over-reliance

³¹¹ Subotnik, *supra* note 2, at 402.

³¹² *Id.* at 342–43.

upon cognitive-based testing is an intellectually lazy approach to determining merit in general and with respect to entry to law school and the legal profession.

Society is best served when we engage in the intellectual and practical work necessary to examine and develop more holistic assessments. Doing so will both reduce the adverse impact that occurs when we rely solely upon cognitive test scores and produce better qualified students, employees and lawyers.