MORE INTELLIGENT DESIGN:
TESTING MEASURES OF MERIT

Kimberly West-Faulcon

This Article articulates the theoretical, legal, and policy implications of new and improved theories of intelligence and recent research finding that conventional mass-marketed standardized tests, or "factorist tests," have less predictive power and larger racial differences in scores than newer multi-dimensional "systems-based" tests. It raises a new question about the fairness of the role that traditional admissions tests like the SAT, GRE, LSAT, and MCAT currently play in selective higher education admissions—whether basing admissions on scores on such tests unfairly distorts the true admissions-related merit of individual applicants and racial groups. The core of this argument is not that selective universities rely on a flawed definition of merit or that traditional factorist tests are racially, economically, or culturally biased. Instead, this Article considers the ramifications of social science evidence suggesting that the admissions tests most commonly relied upon today are less successful in predicting applicants' future academic performance and have more racially skewed scores because they are designed according to a scientifically flawed theory of intelligence. It argues that this analysis is particularly salient in light of recent scientific studies, such as one finding that the currently dominant factorist college admissions test—the SAT—produces significantly larger racial group differences in test scores but with only half the predictive power of a newly designed "more intelligent" college admissions test—a new test based on the theory that intelligence is broadly comprised of more aspects and components than the general intelligence factor "g." Acknowledging that the predictive power of mental tests is inherently limited—even a mental test designed according to a perfectly accurate theory of intelligence could not fully explain differences in individuals' future academic performance, this Article posits that the ramifications of using outmoded and theoretically flawed standardized tests warrant legal examination and necessitate an interdisciplinary endeavor to delineate the framework for such legal analysis. As such, the Article interrogates the prevailing view that comparing applicants' scores on conventional factorist tests affords each applicant the opportunity to be judged fairly according to his or her individual merit. The Article posits that universities that rely on conventional admissions tests like the SAT are potentially vulnerable to the policy criticism that reliance on factorist tests like the SAT is anti-meritocratic—it rewards applicants based on a theoretically flawed and inferior tool for measuring admissions-related merit. In addition, two

* Professor of Law, Loyola Law School, Los Angeles; B.A., Duke University; J.D., Yale Law School. I wish to thank Devon Carbado, Kimberlé Crenshaw, Kent Faulcon, Micheal Gerhardt, Lani Guinier, Cheryl Harris, Luke Harris, Terry Maloney, Alexandra Natapoff, Lee Petherbridge, Valerie Purdie-Vaughns, and Julie Su for their insights, comments, and encouragement. I would also like to thank my research assistants Shawn Chang and Daan Erikson, Joshua Phillips of the Loyola Law School Law Library, and the editors of the University of Pennsylvania Journal of Constitutional Law. Special thanks to my students in Intelligence, Testing, and the Law from Spring 2009 to the present for their broadly diverse perspectives on issues relevant to this project. This Article also benefited from comments made by participants attending faculty workshops at Vanderbilt Law School and the University of North Carolina School of Law, the University of Houston Law Center Reenvisioning Law Colloquium, the Southeastern Association of Law Schools New Scholar Workshop, the National Black Law Journal 35th Anniversary Symposium at UCLA School of Law, the Third Annual Critical Race Studies Symposium at UCLA School of Law, and the June 2009 Colorblind Seminar at the Center for Advanced Study in the Behavioral Sciences at Stanford University.
potential legal implications flow from this Article’s analysis. First, scientific research that reveals that factorist tests have less predictive power but larger racial differences makes universities that rely on such tests vulnerable to Title VI disparate impact administrative complaints. Second, the scientific “test deficiency” of factorist standardized tests provides an alternative or corollary to the diversity rationale for race-consciousness in selective admissions recognized by the Supreme Court’s Fourteenth Amendment Equal Protection jurisprudence—a new “test deficiency” rationale for affirmative action. Correcting for the demonstrated tendency of factorist tests to make individually and racially skewed errors in assessing the admissions-related merit of applicants is meritocratic, fair, legal, and, in certain circumstances, legally required.

**Table of Contents**

**INTRODUCTION** ........................................................................................................... 1237

**I. EXAMINING THE CONFLATION OF HIGH TEST SCORES AND MERIT** ........................................... 1244
   A. Presumed Merit of High Scorers ................................................................. 1247
   B. The Diversity Rationale’s Tacit Conflation .............................................. 1248
   C. Thomas’s Test Deficiency Critique .............................................................. 1251

**II. FROM GENERAL INTELLIGENCE TO SELECTIVE ADMISSIONS** .......................................................... 1255
   A. The Rise of Factorist Tests and America’s Meritocracy
      Ideology ........................................................................................................... 1257
      1. Early Use of the SAT ............................................................................... 1259
      2. The Promise of the SAT’s Predictive Power ............................................. 1261
   B. Utility and Racial Problems of Factorist Tests: The SAT
      Example .......................................................................................................... 1264
      1. Modern Predictive Power of the SAT ....................................................... 1265
      2. Differences in Averaged SAT Scores of Racial Groups .................................. 1270

**III. NEW UNDERSTANDINGS OF INTELLIGENCE:**
   ABOVE AND BEYOND “g” ............................................................................. 1272
   A. Critiques of “g” ............................................................................................... 1273
   B. Examples of Innovation in Intelligence Theory and Testing ....................... 1275
   C. A More Intelligent College Admissions Test? ............................................ 1277

**IV. IMPLICATIONS OF MORE INTELLIGENT TESTS** ................................................................. 1280
   A. Prior Legal Disputes Related to Race, Merit, and Testing ......................... 1281
   B. Resolving the Tension Between Merit and Diversity .................................. 1286
   C. “Test Deficiency” as Complaint and Defense ............................................. 1289
      1. Title VI Test Deficiency .............................................................................. 1291
      2. Reverse Discrimination Equal Protection
INTRODUCTION

Just as new knowledge in metallurgy, electronics, and biochemistry has led to “long-lasting” razor blades, color television, and new medications and vaccines, “psychometrics”—the psychological study of how to measure mental ability—has produced mental tests such as the SAT,^2^ Law School Admissions Test (“LSAT”), Graduate Records Examination (“GRE”), and Medical College Admissions Test (“MCAT”).^3^ Since the turn of the twentieth century, psychometric theories as to how individuals differ in mental ability and how such differences can be measured have impacted legal and lay perceptions of what it means to judge applicants to elite higher education on “individual merit.” This Article asserts that modern social science re-

---


2 Originally called the Scholastic Aptitude Test, “SAT” no longer has an official meaning. TONY MONCHINSKI, CRITICAL PEDAGOGY AND THE EVERYDAY CLASSROOM 171 (2008) (observing that the acronym SAT once stood for “scholastic aptitude test” and later “scholastic assessment test” but that “[q]uestions of what the SAT supposedly assessed led to the jettisoning of that acronym and today the initials SAT stand for nothing”) (emphasis in original). This Article uses “SAT” to refer to the current SAT Reasoning Test, formerly named the SAT I: Reasoning Test. See Frequently Asked Questions, CollegeBoard.com, http://sat.collegeboard.com/about-tests/sat/FAQ (last visited May 11, 2011).

3 Psychologist John Carroll first drew this analogy. See John B. Carroll, The Measurement of Intelligence, in HANDBOOK OF HUMAN INTELLIGENCE 29, 29 (Robert J. Sternberg ed., 1982) (discussing the development of mental tests and the correlation between scores on intelligence quotient (“IQ”) tests and other standardized admissions tests). In using the term “mental tests,” I mean to refer to tests of abstract reasoning. IQ tests, and undergraduate and graduate admissions tests like the SAT, LSAT, GRE, and MCAT. There are substantial similarities and high correlations in scores when “IQ-type” tests and college admissions tests like the SAT are compared.

Most intelligence tests are similar in many ways to tests of academic accomplishment, such as the SAT, and they are often mistaken for one another. This is in part because the two types of tests—intelligence tests and achievement tests—commonly feature time limits and multiple-choice questions. See ANNA T. CIANCIOLO & ROBERT J. STERNBERG, INTELLIGENCE: A BRIEF HISTORY 31 (2004) (“Realistically speaking, all tests of mental ability—whether they are intelligence tests or achievement tests—measure both intellectual aptitude and the outcomes of learning.”).
search focused on how to better define and measure intelligence challenges, the amount of value and reliance placed on conventional versions of such tests, both in terms of how well they measure intelligence, and how accurate they are in predicting success in higher education. This assertion has theoretical, as well as practical, legal, and policy, implications. This Article’s focus is the significance of innovation in intelligence theory and test development over the past few decades that has resulted in the creation of “more intelligent” standardized tests—tests that do a better job of predicting test-takers’ future academic performance—when compared to the conventional educational admissions tests like the SAT, GRE, LSAT, and MCAT.

4 Although the Educational Testing Service (“ETS”) has gone to great lengths to disassociate the admissions tests it designs and administers—the SAT, GRE, LSAT, and MCAT—from IQ tests like the Stanford-Binet, intelligence testing experts consider the SAT and its graduate school-level counterparts to be highly “gloaded.” See, e.g., Richard C. Atkinson & Saul Geiser, Reflections on a Century of College Admissions Tests, 38 EDUC. RESEARCHER 665, 666 (2009) (asserting that “the one constant has been the SAT’s claim to gauge students’ general analytic ability, as distinct from their mastery of specific subject matter”); Linda S. Gottfredson & James Grouse, Validity Versus Utility of Mental Tests: Example of the SAT, 29 J. OF VOCATIONAL BEHAV. 363, 365–66 (1986) (arguing that SAT and other ETS tests are likely good measures of “g”); see also Meredith C. Frey & Douglas K. Detterman, Scholastic Assessment or g?: The Relationship Between the Scholastic Assessment Test and General Cognitive Ability, 15 PSYCHOL. SCI. 373, 377 (2004) (“[T]he SAT is an adequate measure of general intelligence . . . .”).

5 Title VI of the Civil Rights Act of 1964 prohibits federally funded universities from excluding applicants on the basis of race. 42 U.S.C. § 2000d (2006). The current Court has interpreted Title VI itself to prohibit only disparate treatment discrimination. See Kimberly West-Faulcon, The River Runs Dry: When Title VI Trumps State Anti-Affirmative Action Laws, 157 U. PA. L. REV. 1075, 1123 (2009) (“[N]o private right of action exists to enforce Title VI disparate impact regulations.”). The U.S. Department of Education regulations implementing Title VI prohibit federally funded educational institutions from using selection criteria that have an unjustified racially discriminatory effect. Id. at 1122–23 (“A violation of Title VI regulations does not require proof of purposeful discrimination.”). Rejected applicants cannot sue universities for alleged Title VI disparate impact violations. See Alexander v. Sandoval, 532 U.S. 275, 293 (2001) (holding that “no such right of action exists”). However, individuals may file complaints with the U.S. Department of Education Office of Civil Rights (“OCR”) and the U.S. Department of Justice (“DOJ”) alleging such violations. The OCR and DOJ have authority to investigate potential violations of Title VI and to terminate federal funding to educational institutions found to be in violation of Title VI disparate impact regulations. See Investigation Procedures Manual for the Investigation and Resolution of Complaints Alleging Violations of Title VI and Other Nondiscrimination Statutes, U.S. DEP’T OF JUSTICE (Sept. 1998), available at http://www.justice.gov/crt/about/cor/Pubs/manuals/complain.pdf. The DOJ may also file suit against universities to prove Title VI violations. See Title VI Legal Manual, U.S. DEP’T OF JUSTICE, 104–05 (Jan. 11, 2001), available at http://www.justice.gov/crt/about/cor/coord/vimanual.pdf.

6 I have applied a similar analysis to innovations in employment testing. See generally Kimberly West-Faulcon, Fairness Feuds: Title VII’s Competing Conceptions of Discriminatory Testing, 46 WAKE FOREST L. REV. (forthcoming 2011).
that currently dominate selective college and graduate school admissions.

This Article is an interdisciplinary project that considers the implications of technological advances in mental measurement in the context of selective higher education admissions. To do so, it assumes that the goal of American selective universities is to distribute the valuable social and economic good of elite higher education according to a “traditional liberal conception of merit.” Next, it suggests that newly developed intelligence theories and mental tests call into question the conventional wisdom that selecting the highest scorers on conventional factorist tests like the SAT, GRE, LSAT, and MCAT constitutes “meritocratic” selection.

The standardized admissions tests used by most selective universities, like the SAT, GRE, LSAT, and MCAT, are based on the mathematically driven “factorist” view of intelligence. Factorist theories of intelligence conceive a general mental energy—the “g” general intelli-

---

7 The traditional liberal conception of merit I presume throughout is essentially the “liberal equality” conception that applicants should be selected based on a combination of “natural talent” and “conscientious effort.” See JOHN RAWLS, A THEORY OF JUSTICE 73 (1971) (describing and rejecting the liberal interpretation of the second principle of justice as “those who are at the same level of talent and ability, and have the same willingness to use them, should have the same prospects of success regardless of their initial place in the social system, that is, irrespective of the income class into which they are born”). Rawls rejects liberal equality because he concludes that “the initial endowment of natural assets and contingencies of their growth and nurture in early life are arbitrary from a moral point of view.” Id. at 311–12. Justifying or critiquing the liberal conception is beyond the scope of this Article. I will note my agreement with Rawls’s conclusion that “the effort a person is willing to make” as well as their “natural talents” are morally arbitrary: they do not correlate to moral. See id. Here, my focus is assessing whether selective universities’ traditional admissions policies are consistent with that conception. Accordingly, this Article starts from the position that America’s elite public and private universities have adopted a meritocracy ideology that seeks to use objective and unbiased criteria to compare and select applicants.

8 This Article uses the term “factorist” to refer to tests and theories of intelligence derived from defining intelligence according to its structure as revealed by the mathematical technique of “factor analysis.” Factor analysis is a mathematical operation used in a wide variety of academic fields to test or confirm commonalities in data. See, e.g., DENNIS CHILD, THE ESSENTIALS OF FACTOR ANALYSIS (3d ed. 2006). The common attribute of factorist theories of intelligence is the centrality of “g”—the “general intelligence factor.” To “g” theorists, the “g” general factor is “the factor common to all problem-solving abilities.” PAUL KLINE, INTELLIGENCE: THE PSYCHOMETRIC VIEW 3 (1991).

9 In this Article, I use the phrase “factorist tests” and “factorist theories” to refer to psychological tests and theories derived by factor analysis of mental test scores to conclude that intelligence is best conceived as the “g” general factor, a concept first introduced in the early 1900s by the inventor of factor analysis, Charles Spearman. Modern psychometric views of intelligence ascribe to a modernized version of Spearman’s g-based theory. See infra Part II.A.
gence factor”—that is fixed during adulthood and constitutes the most significant aspect of intelligence.

Traditional factorist admissions tests do a better job of predicting which test-takers will succeed academically than would be possible if mental tests did not exist at all; the correlation between scores on such tests and post-admission early grades has been shown to be statistically significant. However, that correlation is not necessarily substantively significant. Even though factorist tests do provide some predictive information, they also leave a lot of the differences in test-takers’ future academic success in college and graduate school unexplained. This means that the overall predictive power of factorist tests leaves substantial room for innovation and improvement.

The focus of this Article is new intelligence research that posits that flaws in the g-based theory of intelligence may explain why g-oriented tests are useful but still weak predictors of future academic performance. Theorists who have posited alternatives to the g-based

10 This is the scientific theory that intelligence is equated to a single “general ability” called the “general intelligence factor” or “general factor” ("g") that is common to solving problems on all tests of mental ability. See Arthur R. Jensen, Spearman’s g: Links Between Psychometrics and Biology, in BRAIN MECHANISMS: PAPERS IN MEMORY OF ROBERT THOMPSON 103, 103 (Francis M. Grinella & Jen Yu eds., 1993) (“g refers to the component of individual differences variance that is common to all tests of mental ability.”). g has also been defined as the ability to take mental tests. See, e.g., KLINE supra note 8, at 3–4 (“Since the best intelligence tests are deliberately constructed to measure this g factor it does make sense to define intelligence as what intelligence tests measure, provided that the g factor can be specified.”).

11 See Richard Lempert, The Significance of Statistical Significance: Two Authors Restate an Incon- trovertible Caution. Why A Book?, 34 L. & SOC. INQUIRY 225, 226 (2009) (“The caution: Statistical significance is not the same as substantive significance. Statistically significant relationships may, and often do, tell us nothing that matters, while relationships that do not achieve conventional levels of statistical significance can be important, and we may neglect them at our peril.”).

12 See id.

13 If factorist test scores were the only information available about a set of applicants, admission based on test scores would result in the selection of more academically successful applicants than random selection. See infra Part II.B.

14 Modern intelligence research suggests factorist tests are like the traditional head-on frontal collision tests adopted by the National Highway Traffic Safety Administration (“NHTSA”) nearly thirty years ago; they only partially explain the differences in future outcomes society wants explained as compared to new and more useful tests. See Ronald Montoya, NHTSA Revises Five-Star Safety Ratings, EDMUND’S.COM (Oct. 1, 2010), http://www.edmunds.com/car-safety/nhtsa-revises-five-star-safety-ratings.phtml (describing the inclusion of the new tests as “one of the biggest changes since safety testing was first conducted in 1978” along with a recognition of models with new crash avoidance technology features, such as electronic stability control, lane-departure warning systems
theory of intelligence and alternatives to \textit{g}-based tests contend their new “multi-dimensional” approaches to defining intelligence and mental measurement are more predictive of test-takers’ future academic performance than traditional \textit{g}-based factorist tests.\textsuperscript{16} This Article considers whether this particular research calls into question the fairness of the role that factorist tests currently play in admission to America’s most selective colleges and universities.

Although the central examination of this Article—the implications of scientific innovation in the capacity of mental tests to measure test-takers’ admissions-related merit—is not about race or race discrimination, the modern intelligence research that is its focus has racial implications. Numerous studies show that newly designed tests are not just more valid and predictive than traditional factorist tests, but this research also suggests that differences in racial group average scores are smaller on tests based on more complete theories of intelligence (multi-dimensional conceptions of intelligence) than on factorist tests. Such findings—essentially that at least some portion of the racial differences in factorist test scores are attributable to inadequacies in factorist tests as predictors of future academic success—have significant legal and policy implications related to race and selective higher education admissions.

Specifically, scientific evidence that “more intelligent” tests have \textit{smaller} racial differences in scores shifts the legal terrain in conflicts involving race and testing. Research demonstrates that the predictive inadequacies of, and scientifically unjustified racial differences in, scores on conventional factorist tests like the SAT may be legally cognizable “test deficiencies.”\textsuperscript{17} Here, I suggest that Fourteenth Amendment equal protection jurisprudence and Title VI disparate impact standards are structured to recognize the concept of test deficiency that is the focus of this Article. Scientific evidence that factorist admissions tests have larger racial differences in scores than newly designed tests could be relied upon to demonstrate this type of test deficiency, making institutions that rely on factorist tests more vulnerable to allegations from minority applicants that their rejection was anti-meritocratic and illegal, and creating a new legal defense to

\begin{itemize}
\item \textsuperscript{16} See infra Part III.
\item \textsuperscript{17} See West-Faulcon, \textit{supra} note 5, at 1122-23 (describing potential liability for discrimination under Title VI for using “selection criteria in a manner that constitutes effect discrimination”).
\end{itemize}
the “reverse discrimination” version of such allegations\textsuperscript{18} by rejected whites—this raises the prospect of test deficiency supporting Title VI disparate impact complaints\textsuperscript{19} against selective universities and a new legal defense of race-based affirmative action under the Fourteenth Amendment.\textsuperscript{20}

Part I describes the dilemma presented by the approach that universities have used tacitly to defend race-based affirmative action in higher education admissions—that race consciousness is needed to reconcile the purported inherent tension between merit-based selection and racial diversity. This Part articulates the need for more intelligent tests—tests that better measure admissions-related merit with smaller racial gaps in test scores—that can potentially resolve this dilemma. It also contrasts the frequency with which reliance on non-test, non-grade admissions variables are challenged as deviations from merit with this Article’s consideration of whether reliance on scores from tests based on scientifically flawed theories of intelligence deprive applicants of fair assessment on the basis of the traditional conception of higher education admissions-related merit.

Part II of this Article provides historical background on the origins of the concept of “g”—the “general factor” of intelligence—and factorist theories of intelligence as well as early and modern critiques of the conception of intelligence as “g.” It explains the salient role that the technology of mental testing, specifically, factorist admissions tests like the SATs, have traditionally played in implementing the American ideology of meritocracy in the context of selective higher education admissions. This Part also explicates how, prior to the existence of new intelligence research highlighted in this Article, it was generally assumed that, due to their predictive power, factorist tests like the SAT were the most objective and scientific proxies available

\textsuperscript{18} The Court essentially uses the same analysis to evaluate Fourteenth Amendment reverse discrimination claims filed against public universities as it does to evaluate reverse discrimination filed against private selective universities under Title VI. See Alexander v. Choate, 469 U.S. 287, 293–95 & n.11 (1985) (explaining Guardians Ass’n v. Civil Serv. Comm’n of N.Y., 463 U.S. 582 (1983)). In Guardians Ass’n, a majority of the Court held that a violation of Title VI required proof of discriminatory purpose, while a different majority held that proof of discriminatory effect suffices when the suit is brought to enforce regulations issued pursuant to Title VI. Guardians Ass’n, 463 U.S. at 608 n.1 (Powell, J., concurring) (detailing the multiple holdings of the Court). Liability under Title VI itself is identical to the Fourteenth Amendment Equal Protection Clause in its requirement that plaintiffs prove discriminatory intent. See Washington v. Davis, 426 U.S. 229, 240 (1976).

\textsuperscript{19} Title VI disparate impact complaints may be filed with the United States Department of Education Office of Civil Rights. See supra note 5.

\textsuperscript{20} See infra Part IV.
for assessing an applicant’s admissions-related merit. Next, using the SAT as an example, it describes both the limitations in the predictive capacity of factorist tests and differences in racial groups’ averaged scores on conventional factorist tests.

Part III examines contemporary theories of intelligence, referred to in this Article as “multi-dimensional” theories of intelligence. It focuses specifically on the operationalization of American psychologist Robert Sternberg’s Triarchic Theory of Successful Intelligence as new triarchic college admissions tests of “creative,” “practical,” and “analytic” intelligence. It describes a 2006 research project that revealed that triarchic admissions tests based on a systems-based theory of intelligence were twice as predictive of freshman college grades as the g-loaded SAT with smaller racial differences in test scores.

Part IV of this Article articulates the theoretical, legal, and policy implications of improved theories of intelligence and empirical evidence of the enhanced predictive power of tests designed according to newer theories of intelligence. In addition, it asserts that research demonstrating that factorist tests have, first, less predictive power and, second, larger racial differences in scores than are scientifically justified—two major “test deficiencies” make universities that rely on such tests vulnerable to Title VI disparate impact complaints. This Part draws from Justice Powell’s decision in Bakke to articulate how such test deficiencies may form the basis for a legal justification for greater reliance on systems-based tests, non-test score criteria, or for considering race in admissions to correct for unjustified racial skews in factorist tests results—racial differences in scores that stem from the theoretical deficiencies of factorist theories of intelligence and the predictive deficiencies of conventional factorist tests.

The Article concludes that scientific intelligence research of modern multi-dimensional theories of intelligence, such as systems-based intelligence theories, has significant implications. Among the most significant is its finding that college admissions tests designed accord-


\[22\] Id. (“Based on multiple regression analyses, for our sample, the . . . [systems-based] measures alone approximately double the predicted amount of variance in college GPA when compared with the SAT alone (comparative R values of .199 [explains 19.9% of variation in college GPA] to .098 [explains 9.8% of variation in college GPA], respectively.”).

\[23\] See West-Faulcon, supra note 5, at 1110 (describing test-deficiency theories).
ing to newer theories of intelligence are more useful admissions tools than conventional tests. If proven true, this research suggests that selection based on rank-order scores on conventional factorist tests like the SAT, GRE, LSAT, and MCAT is anti-meritocratic. In addition, the Article posits that rejected applicants from lower-scoring racial groups, including African Americans and Latinos, may rely on this type of intelligence research to argue that overreliance on factorist tests as admissions criteria is unfair, anti-meritocratic, and illegal under Title VI civil rights disparate impact regulations. In addition, research illuminating theoretical and predictive failings of factorist tests may also impact reverse discrimination claims filed by rejected white applicants claiming an institution’s use of race-based affirmative action in admissions violates the Fourteenth Amendment Equal Protection Clause and Title VI’s prohibition against purposeful race discrimination.

I. EXAMINING THE CONFLATION OF HIGH TEST SCORES AND MERIT

Certainly the tests do seem to do better than chance. But they do not have the value that their deceptively precise scoring system suggests. The proponents’ own data show that, for example, most of those scoring in the bottom 20% on the test do better than that in law school—indeed six of every 100 of them will be in the top 20% of their law school class. And no one knows how many of those who were not admitted because of their test scores would in fact have done well were they given the chance.


It is generally the case that an applicant with a higher factorist test score is viewed as more deserving of admission than a similarly credentialed applicant with a lower score on the same test. As a result,

---

25 Challenges are both formal and informal, ranging from phone calls and letters to admissions offices to complaints with the U.S. Department of Education and lawsuits filed in federal court. For example, a class of “minorities and women qualified for graduate school admission” filed an OCR complaint accusing the University of California Berkeley (“UC Berkeley”) of violating Title VI and Title IX disparate impact regulations after eliminating race- and gender-based affirmative action pursuant to UC Regents Resolution SP-1. Letter from Abby Leibman, Cal. Women’s Law Ctr. et al., to Stefan Rosenzweig, Reg’l Dir., Office of Civil Rights, Region IX, U.S. Dep’t of Educ. (Mar. 19, 1997) (on file with author); see also Complaint at 3, Rios v. Regents of the Univ. of Cal., No. 99-0525 (N.D. Cal. 1999), settled sub nom. Castaneda v. Regents of the Univ. of Cal., No. 99-0525 (N.D. Cal. 2003) (“This is an action challenging the discriminatory failure of Defendants to give full and fair consideration to applications for undergraduate admission to the University of California at Berkeley . . . .”)
university leaders, faculty members, and the general public decry the admission of lower-scoring applicants to the exclusion of higher-scoring ones. Many reach these conclusions based on an essentially “folk” understanding of the nature of mental ability and mental testing.

It is because of this general presumption that rank-ordering by test score aligns with rank-order admissions merit that universities’ reliance on non-test score, non-grade admissions criteria is assumed by many to be a deviation from a true academic merit-based standard. Confidence in the merit-measuring capacity of conventional standardized tests like the SAT is so great that universities like the University of California, Berkeley (“UC Berkeley”) and the University of California, Los Angeles (“UCLA”) that consider selective non-test admissions criteria, such as whether an applicant has faced adversity, are often accused of presiding over a “best sob-story” sweepstakes and of granting illegal “preferences” to minority applicants. The accusations that California’s most selective public universities have been “cheating”—violating state anti-affirmative action laws—are premised

---

26 Most lay and legal discourse adopts what could be described as a “folk definition” of intelligence that aligns fairly closely to conventional unitary and hierarchical g-centered theories of intelligence. Cf. ROGERS ELLIOT, LITIGATING INTELLIGENCE: IQ TESTS SPECIAL EDUCATION, AND SOCIAL SCIENCE IN THE COURTROOM 7 (1987) (citing Lloyd G. Humphreys, General Intelligence, in PERSPECTIVES ON BIAS IN MENTAL TESTING (Cecil R. Reynolds and Robert T. Brown eds., 1984)) (describing the “folk definition” of intelligence as “an innate, fixed and measurable capacity utilized to varying degrees depending upon the adequacy of an individual’s environment and the effort the individual expends”).

27 America’s most elite private universities have long considered non-numerical, qualitative information—non-test score criteria—to augment applicants’ SAT scores and other quantitative admissions data such as high school GPA. See generally JEROME KARABEL, THE CHOSEN: THE HIDDEN HISTORY OF ADMISSION AND EXCLUSION AT HARVARD, YALE, AND PRINCETON 44 (2005) (discussing admissions to Harvard, Yale, and Princeton over the past century). In recent years, many of the nation’s most selective public universities have followed suit—adopting non-numerical, “holistic” policies that rely on various non-test score factors to make admissions decisions. For public universities that consider race as a factor in admissions, the Supreme Court has held that the Equal Protection Clause requires their admissions policies be non-numerical and holistic. See Gratz v. Bollinger, 539 U.S. 244, 275 (2003) (finding a policy not “narrowly tailored” because the policy did not contain “individualized consideration”); Grutter v. Bollinger, 539 U.S. 306, 337 (2003) (“The importance of this individualized consideration in the context of a race-conscious admissions program is paramount.”).

28 This shift has been, for many top-ranked public universities, prompted by the Supreme Court’s holding in Grutter that the use of race-based affirmative action is only narrowly tailored if it is part of a holistic, as opposed to numbers-driven, admissions process. Grutter, 539 U.S. at 343.

29 See West-Faulcon, supra note 5, at 1086–95 (describing the origins of state anti-affirmative action laws and how such laws prohibit public universities from granting preferential treatment on the basis of race).
implicitly on the assumption that admitting students inconsistently with the rank order of their SAT score is unfair and anti-meritocratic.30

Despite the fact that elite educational institutions are forthright in declaring that they rarely, if ever, admit applicants based solely on scores on admissions test like the SAT and LSAT,31 the institutions rarely offer a coherent explanation of how much they do rely on test scores and why they choose a particular degree of reliance. As opposed to articulating an academic merit-based rationale, admissions officials and university leaders usually defend decisions to consider criteria other than standardized test scores as necessary to fulfill a broad institutional “mission” to educate a “diverse” student body. The implication is that deviation from admission by test score rank order is the price that must be paid in order to admit members of lower-scoring racial groups—a presumptive merit versus diversity trade-off.

It is rare when institutions point to the predictive limitations of traditional factorist tests like the SAT and LSAT or to potential flaws in the theories of intelligence underpinning such tests as the rationale for relying on non-test score criteria. Instead, elite institutions tout their institution’s average SAT, LSAT, GRE, and MCAT as a marker of institutional and student “quality” and “academic selectivity.”32 By doing so, selective universities have made themselves susceptible to criticisms that rejections of higher-scoring applicants is unfair and, if related to race, illegal. A prime example of such criticism is the one lodged by John Moores in 2003 when he contended that UC Berkeley unfairly denied admissions to applicants with very high SAT scores.


32 See generally Grutter, 539 U.S. at 340 (equating student “quality” with test scores, and noting that “decreasing the emphasis for all applicants on undergraduate GPA and LSAT scores . . . would require a dramatic sacrifice of . . . the academic quality of all admitted students” (citation omitted) (internal quotation marks omitted)).

33 Id.
A. Presumed Merit of High Scorers

Former Regent of the University of California John Moores is among those who have criticized public universities for passing over high-scoring applicants in favor of individuals with lower test scores. In 2003, subsequent to the implementation of a new admissions policy at UC Berkeley, then-Regent Moores very publicly criticized the denial of admission to high school students with high SAT scores as unfair. The assumption underlying his assertion was that the consideration of admissions criteria other than grades and test scores had resulted in a dilution of academic merit standards—Moores’s claim was essentially that UC Berkeley had adopted a new admissions policy that permitted the selective admissions equivalent of “line jumping”—more qualified students were unjustifiably being passed over by less qualified students. To support his claim, Moores pointed to UC Berkeley’s denial of 3200 students with SAT scores above 1400 and its decision to admit 386 students with SAT scores of 1000 or below.

Although UC Berkeley did attempt to defend the admissions decisions attacked by Moores, universities across the nation rarely offer an explicit theoretical and empirical rationale for augmenting applicants’ test scores and grades with quantitative information about applicants. This failure to justify admissions criteria and their weighting has made universities vulnerable to political and legal claims that their admissions policies unfairly or illegally fail to select applicants with the greatest admissions-related merit. An additional consequence of this failing is that it reinforces the perception, articulated

34 In 2003, the same year the Supreme Court decided the Grutter case, University of California Regent Moores issued a report on the admissions process at UC Berkeley—California’s most selective public university—examining the number of applicants with high SAT scores denied admission to UC Berkeley. Moores, supra note 135, at 3, 183–214. Moores’s challenge of the denial of admission of thousands of high-scoring applicants to UC Berkeley is an example of the inquiry that has previously been the central focus of legal scholarship—whether it is fair for universities to admit applicants with lower scores on tests like the SAT and reject applicants with higher test scores whose grades may be equivalent to or better than the low-scoring admitted student.

and ascribed to by Justice Thomas in his dissent in *Grutter v. Bollinger*, that the entire process of selective admissions at the nation’s universities is “‘poisoned’ by numerous exceptions to ‘merit.’” Justice Thomas’s critique of the University of Michigan Law School’s admissions policy, like Regent Moores’ critique of UC Berkeley’s undergraduate admissions policy, is emblematic of a common perception that elite public and private educational institutions regularly deviate from selecting individuals based on admissions-related merit.

**B. The Diversity Rationale’s Tacit Conflation**

In *Grutter* and its companion case *Gratz v. Bollinger*, the Supreme Court held that a university’s educational interest in “diversity”—the goal of admitting more than a token number of members of a racial group that would otherwise be present in only very small numbers—constitutes a compelling justification for considering race as an admissions factor. Writing for the majority in *Grutter*, Justice O’Connor concludes that the University of Michigan Law School established that its means of considering race in admissions—as part of a holistic, non-numerical consideration of the applicants’ entire file—was “narrowly tailored” but holds that the more numbers-driven undergraduate admissions policy challenged in *Gratz* is not narrowly tailored.

Central to the Court’s analysis in *Grutter* is its acceptance of the University of Michigan’s empirical claim that considering race as a factor in admissions is the only “workable” means for the law school to admit a critical mass of underrepresented minority applicants without lowering its current academic admissions standards. The Court ultimately distinguishes the Michigan Law School and undergraduate policies based on how rigidly and numerically race was considered as a factor; it accepts the narrow tailoring argument offered

---

37 *Grutter*, 539 U.S. at 568 (Thomas, J., concurring in part and dissenting in part).
38 *Id.* at 329 (majority opinion) (discussing a university’s right to select a diverse student body in seeking to achieve goals important to its mission). The current “diversity rationale” for race-based affirmative action in selective higher education admissions stems from Justice Powell’s dispositive concurring opinion in the case *Regents of the University of California v. Bakke*, 438 U.S. 265 (1978).
39 See, e.g., Ian Ayres & Sydney Foster, *Don’t Tell, Don’t Ask: Narrow Tailoring After Grutter and Gratz*, 85 TEX. L. REV. 517, 518 (2007) (arguing that the Court’s “individualized consideration” requirement encourages government decision-makers to not disclose how much racial preference it is giving).
40 *Grutter*, 539 U.S. at 539–40.
41 *Id.* (“Narrow tailoring does not require exhaustion of every conceivable race-neutral alternative.”).
by the University of Michigan that the institution could not maintain high academic test and grade merit standards without considering applicants’ race as a factor in admissions.\footnote{See Brief for Respondent at 13, Grutter v. Bollinger 539 U.S. 306 (2003) (No. 02-241), 2003 WL 402236 (“No honestly colorblind alternative could produce educationally meaningful racial diversity at present without substantially abandoning reliance on traditional academic criteria, and hence abandoning academic excellence as well.”).}

Although some amici in \textit{Grutter} challenged the presumption that LSAT scores were accurate measures of merit,\footnote{See, e.g., Brief for New York State Black and Puerto Rican Legislative Caucus as Amicus Curiae Supporting Respondents at 2, Grutter v. Bollinger, 539 U.S. 306 (2003) (No. 02-241), 2003 WL 554403. (“Test scores can also be adversely affected by candidates’ unconscious reaction to widespread stereotypes disparaging the intellectual abilities of minority group members.”).} the University of Michigan Law School did not emphasize the predictive limitations of the LSAT. University defendants in reverse discrimination cases rarely argue that they select lower-scoring non-white applicants because they are actually more qualified than higher-scoring whites.\footnote{\textit{Cf.} Charles R. Lawrence, Essay, \textit{Two Views of the River: A Critique of the Liberal Defense of Affirmative Action}, 101 COLUM. L. REV. 958 (2001) (criticizing universities and the “liberal defense . . . of affirmative action” for this failure). \textit{See also} West-Faulcon, \textit{supra} note 5, at 1148 (“[W]hen selective universities invoke the diversity rationale, those institutions are rarely called upon to identify explicitly the characteristics that qualify or disqualify minority applicants.”).} Instead, universities usually take the position that considering race in admissions is the most narrowly tailored means of achieving the educationally optimal level of racial diversity because race-neutral policies would come at too great an expense to their institutions’ high academic standards.

According to the University of Michigan’s position in \textit{Grutter}, race-conscious affirmative action is necessary because so few African-American and Latino students have LSAT scores in the ultra-high ranges—the LSAT range of the most selective law schools in the nation—that the small number of African-American and Latino students with LSAT scores in that highest range will be admitted to multiple top law schools and hence will not all choose to attend Michigan Law School.\footnote{Brief for Respondent, \textit{supra} note 42, at 5 (noting that “[i]n 1997 when petitioner applied, there were only 67 minority [African-American, Latino and Native-American] applicants, compared to 1236 white and Asian American applicants, in the LSAT range (164+) from which over 90% of the admitted white students were drawn”).} In short, the University of Michigan argues that there simply are not enough ultra-high-scoring African-American and Latino applicants to admit significant numbers of them without race-consciousness.\footnote{\textit{Ib.} at 13 (arguing that “[g]iven the . . . population of college graduates, however, law schools like Michigan cannot admit those [minority students whose admissions would in-}
of Michigan adopts the typical approach of universities defending their affirmative action policies under the diversity rationale—asserting that the consideration of race is narrowly tailored because elite educational institutions cannot maintain their desired merit-based academic standards and also admit more than a token number of applicants from lower-scoring racial groups.

This approach to defending affirmative action tacitly conflates high scores on the LSAT with admissions-related merit. Even as the University of Michigan describes the African-American and Latino applicant pool as “talented” and “well-qualified,” 47 LSAT scores are elevated over college grades as a more critical criterion for maintaining high academic admissions standards. 48 Without explaining whether students with ultra-high college grade point averages (“GPAs”) could potentially be as qualified as, or possibly more qualified than, applicants with ultra-high LSAT scores, the University of Michigan implies that there is necessarily a tradeoff that elite law schools make when they admit lower-scoring non-white students.

By stating that “[i]n 2000, there were only 26 African-American applicants nationwide with at least a 3.5 grade point average and a 165 on the LSAT compared to 3173 whites and Asian Americans,” 49 the University underscores how very important it deems LSAT scores despite having acknowledged that an applicant’s LSAT score is an imperfect predictor of future grades in law school. 50 Moreover, in Grutter, the University of Michigan takes no position as to the comparative admissions-related merit of the minority students with LSAT scores below 164 or 165 that it admits under affirmative action. Specifically, the University declines to assert explicitly that some applicants with lower LSAT scores are, in certain instances, more qualified than applicants with higher test scores. 51 Interestingly, Justice Tho-
mas, in dissent, makes this assertion, in part, a basis for concluding that the Michigan Law School’s admissions policy is unconstitutional.

C. Thomas’s Test Deficiency Critique

In his dissent in *Grutter*, Justice Thomas contends that the major problem with Justice O’Connor’s majority opinion is that it accepts the University of Michigan Law School’s inaccurate characterization of its compelling interest in considering race as a factor in admissions. Thomas expresses a great deal of skepticism as to whether it is necessary for Michigan Law School to base admission on LSAT scores. Thomas claims that, contrary to the law school’s assertion that its reason for considering race is to provide students the educational benefits that flow from a racially diverse student body, the law school actually seeks “to improve marginally the education it offers without sacrificing too much of its exclusivity and elite status.” The solution that Justice Thomas proposes is that the defendant law school end its reliance on the LSAT admissions test. Thomas is of the view that the institution could adopt different admissions methods “such as admitting all students who meet minimum qualifications” instead of students with the highest LSAT scores and achieve its goal of racial diversity without needing to rely on race-based affirmative action. In fact, he questions whether standardized admis-
sions tests truly have the capacity to predict who will succeed in law school. Thomas seems to believe African-American students, in particular, would be better served if assessed based on non-test criteria—by “looking for those students who, despite a lower LSAT score or undergraduate GPA, will succeed in the study of law” and then “matching” African-American students to the law school where they are likely to succeed academically.

Having found more wrong than right with the University of Michigan’s selective admissions based on standardized tests, Thomas offers additional reasons to be critical of the impact that reliance on LSAT scores has on lower-scoring groups like African-American students. Beyond questioning the predictive capacity of the LSAT and other standardized admissions tests, Thomas accuses elite universities of using “intelligence tests,” in the past, to intentionally discriminate against applicants on the basis of their race, ethnicity and religion, and of currently using intelligence tests “with full knowledge of their disparate impact” against groups like African Americans. Thomas ultimately exploits the University of Michigan’s tacit suggestion that high academic standards and racial diversity in admissions are in tension by asserting that, to the extent such a tension exists, it can be resolved without race-consciousness; according to Justice Thomas, the university has the option of choosing between diversity and elite-ness. To explain how the University of Michigan could exercise its options, Thomas suggests the adoption of a non-test-driven “certifi-

58 Id.
59 Id. at 372.
60 Id.
61 Id. at 369. On this specific point, Justice Thomas wrote:
The initial driving force for the relocation of the selective function from the high school to the universities was the same desire to select racial winners and losers that the Law School exhibits today. . . . Columbia employed intelligence tests precisely because Jewish applicants, who were predominantly immigrants, scored worse on such tests. Thus, Columbia could claim (falsely) that “[w]e have not eliminated boys because they were Jews and do not propose to do so. We have honestly attempted to eliminate the lowest grade of applicant [through the use of intelligence testing] and it turns out that a good many of the low grade men are New York City Jews.” In other words, the tests were adopted with full knowledge of their disparate impact.

Id. (citations omitted).
62 Justice Thomas asserts that top ranked public graduate schools, colleges and universities like the University of Michigan and its law school do not have a compelling interest in using race-conscious admissions because they have the option of choosing whether to be racially “diverse” or “elite.” See id. at 372 & n.11, 373 (discussing the subjective assessments of “know-it-all-elites”). Thomas suggests public higher education should not be elite—it should be available to any “certified” graduate who completes a required course of study. Id. at 368–69.
cate system” admissions process. Presumably, under this system of certification, students who complete a particular course of undergraduate study would be admitted without reliance on standardized test scores and then retained or expelled from the law school according to that individual’s actual performance as a law student.

Thomas’s suggestion highlights the relationship between selective universities’ reliance on standardized tests and their institutional prestige. Doctrinally, Thomas argues that Michigan Law School’s failure to adopt race-neutral alternatives like the certificate system of admissions to achieve the benefits of racial diversity demonstrates that the challenged race-conscious admissions policy should be deemed unconstitutional due to the existence of race-neutral alternatives. Moreover, Thomas does not believe that Michigan Law School’s reliance on the LSAT serves any real educational purpose; he thinks the law school uses the LSAT to admit ultra-high-scorers on the LSAT because that translates into “selectivity” which is a marker of elite-ness. Yet, he vacillates between criticizing the LSAT—on one hand he expresses his view that the LSAT is a weak predictor of success in law school and that the racial gap in scores could be closed if African Americans had greater incentive to invest in more test preparation—and acquiescing to the LSAT’s assessment of African-American students—he thinks the racial gap in LSAT scores will still exist in twenty-five years.

Ultimately, Justice Thomas’s quandary suggests a vulnerability for selective universities that differs from the reverse discrimination-type claim made on behalf of rejected whites in the Grutter case. Innovations in testing technology—the development of standardized tests that more accurately predict test-takers’ future academic success with less racially skewed scores than traditional admissions tests like the LSAT and SAT—increase the likelihood that rejected non-whites will challenge their rejection based on conventional standardized admissions tests. Even though tests like the LSAT and SAT are suffi-

---

63 Id. at 368–69.
64 Id.
65 See West-Faulcon, supra note 5, at 1108 (discussing how the average SAT score is used to gauge an institution’s prestige).
66 Grutter, 539 U.S. at 568-69 (Thomas, J., concurring in part and dissenting in part).
67 I have articulated elsewhere that the capacity of standardized tests as “prestige-enhancing” tools may be more significant to elite educational institutions than their merit-measuring capacity. See, e.g., West-Faulcon, supra note 5, at 1083.
68 Grutter, 539 U.S. at 367 (Thomas, J., concurring in part and dissenting in part).
69 Id. at 363.
70 See infra Part III.
ciently predictive for universities to legally justify reliance on them in the absence of equally predictive and less racially skewed tests, modern intelligence theories leading to the development of more predictive tests with smaller racial differences in scores make reliance on traditional tests more difficult to justify. At their current levels of predictive power—explaining less than 20% of the variation in test-takers’ future academic performance, \(^{71}\) conventional standardized admissions tests are sufficiently imperfect that institutions placing inappropriately heavily reliance on them are potentially vulnerable to policy critiques and legal challenges on “test deficiency” grounds. It is my contention in this Article that test deficiency-type claims, if empirically verifiable, may be asserted by rejected non-white applicants filing Title VI disparate impact administrative complaints.

While it is unclear whether the current Court would entertain test deficiency as a justification for race-based affirmative action in admissions, earlier Justices have. Most notably, Justice Powell’s controlling opinion in *Regents of the University of California v. Bakke\(^ {72}\)* suggests an analytic framework for considering the deficiencies of standardized tests as a justification for reduced reliance on test scores or race-consciousness to compensate for such tests’ failings. \(^{73}\) Dissenting in *DeFunis v. Odegaard*, \(^{74}\) Justice Douglas seems willing to apply distinct legal analysis when scores on the LSAT test misclassify test-takers’ admissions-related merit due to test deficiencies. \(^{75}\) Below, this Article

---

\(^{71}\) See infra Part II.A.


\(^{73}\) *Id.* at 306 n.43 (concluding that the consideration of race as a factor in medical school admissions is arguably not racially preferential if adopted to cure “established inaccuracies” in the capacity of the MCAT medical school admissions test to predict academic performance).


\(^{75}\) In his dissent to the per curiam opinion dismissing a case filed by a white applicant to the University of Washington claiming the law school’s affirmative action policy violated his rights under the Fourteenth Amendment’s Equal Protection Clause, Justice Douglas drew attention to the defendant law school’s failure to argue that minority applicants with lower LSAT scores might nevertheless have greater admissions-related merit than white applicants with higher LSAT scores:

> [B]y whatever techniques, the law school must make choices. Neither party has challenged the validity of the [LSAT-GPA composite score] Average employed here as an admissions tool, and therefore consideration of its possible deficiencies is not presented as an issue. The Law School presented no evidence to show that adjustments in the process employed were used in order validly to compare applicants of different races; instead . . . [the Law School] chose to avoid making such comparisons . . . . To the contrary, the school appears to have conceded that by its own assessment—taking all factors into account—it admitted minority applicants who would have been rejected had they been white.
examines the significance of new scientific research on intelligence and mental testing that did not exist when Justice Powell and Justice Douglas expressed their seeming willingness to entertain test deficiency defenses to reverse discrimination lawsuits filed by rejected white applicants. To further explicate the "test deficiency" framework, Parts II and III respectively, describe, first, the origins and current uses of conventional standardized admissions tests by focusing on the SAT as an example and, second, how modern multidimensional theories of intelligence challenge the conventional theories of intelligence upon which tests like the SAT are modeled.

II. FROM GENERAL INTELLIGENCE TO SELECTIVE ADMISSIONS

Test constructors will continue to employ factorial procedures, provided they pay off in improving the efficiency and predictive value of our test batteries . . . . The continuous difficulties with factor analysis over the last half century suggest that there may be something fundamentally wrong with models which conceptualize intelligence in terms of a finite number of linear dimensions.

—Read D. Tuddenham

While philosophical discussions about the nature of human intelligence date back thousands of years, the first scientific theory of in-

---

Id. at 330–31. Justice Douglas took it upon himself to make the argument, which the law school did not, by asserting the limitations in predictive capacity of admissions test like the LSAT:

Of course, the law school that admits only those with the highest test scores finds that on the average they do much better, and thus the test is a convenient tool for the admissions committee. The price is paid by the able student who for unknown reasons did not achieve that high score—perhaps even the minority with a different cultural background. Some tests, at least in the past, have been aimed at eliminating Jews.

Id. at 329. (discussing data showing that six of every 100 students scoring in the bottom 20% of the LSAT end up in the top 20% of their law school class). This Article’s focus is on the theoretical and predictive inadequacies of conventional factorist tests, not allegations that such tests are culturally biased. By contrast, Justices Powell and Douglas seem to assume that the test deficiency would be cultural bias.


77 See, e.g., IMMANUEL KANT, CRITIQUE OF JUDGEMENT 146 (Nicholas Walker ed., James Creed Meredith trans., 2d ed. 2007) (“Genius, according to these presuppositions, is the exemplary originality of the natural endowments of a subject in the free employment of his cognitive faculties.”); PLATO, REPUBLIC 264 (Robin Waterfield trans., 1993) (“The summit of the intelligible realm is reached when, by means of dialectic and without relying on anything perceptible, a person perseveres in using rational argument to approach the true reality of things until he has grasped with his intellect the reality of goodness itself.”); see also CIFICIOLO & STERNBERG, supra note 3, at 2 (“[I]deas about the nature of intelligence have existed for thousands of years . . . .”).
intelligence was introduced at the turn of the twentieth century. In an article written in 1904, British psychologist Charles Spearman posited, based on the application of a mathematical operation Spearman invented called “factor analysis,” that the reason scores on intelligence tests are correlated with each other is attributable to a dominant “general factor” that governs all mental ability. Spearman deemed “g” as that general factor of intelligence because he believed it to be a mental ability that is common to all subjects and therefore distinguishable and more significant than what Spearman termed “specific” mental abilities—those specific to particular subject matter. The hallmark of Spearman’s g-based theory of intelligence or the g-oriented psychometric perspective is that what is common across all intelligence tests is a single, unitary, linearly rankable, measurable generalized mental energy.

However, dating back to the earliest days of mental testing, g-based theories of intelligence have been challenged and alternative theories posited. Experts have also theorized that intelligence is “information processing”—accurately identifying trends and patterns,
speed, including reaction and inspection time,\textsuperscript{86} one’s ability to adapt,\textsuperscript{87} working memory, and application of memory.\textsuperscript{88} Some have conceived of intelligence as prior knowledge, such as expertise in language and vocabulary\textsuperscript{89} and multiplicities of intelligence.\textsuperscript{90} Hence, differing perspectives as to how intelligence should be conceptualized are as old as the concept of “g.” Over the past thirty years, intelligence research challenging the theoretical underpinnings of g-oriented factorist tests has further spurred the debate over whether a “general factor” of intelligence or singular “g” exists.\textsuperscript{91}

A. The Rise of Factorist Tests and America’s Meritocracy Ideology

Almost contemporaneously with Spearman’s development of a g-centered theory of intelligence in the early 1900s, French psychologist Alfred Binet developed a new tool for measuring mental ability—a new mental test that proved to be as influential as Spearman’s theory. Binet’s early “intelligence test” was designed to diagnose children with learning problems for placement in a “special education” program.\textsuperscript{92} The Binet test was administered one-on-one by a

\begin{itemize}
  \item See Arthur Jensen, Clocking the Mind: Mental Chronometry and Individual Differences 244 (2006) (describing “chronometry,” or the measurement of cognitive speed, as a “uniquely valuable instrument for measuring interindividual and intrapersonal variation in many cognitive phenomena”); C. Spearman, The Abilities of Man: Their Nature and Measurement 245 (1927) ("If we desire any genuine measurement of cognitive ability, it is to these universal quantitative properties of clearness and speed that we are obliged to turn.").
  \item See Alfred Binet & Theodore Simon, The Development of Intelligence in Children 43 (Elizabeth S. Kite trans., 1916) (“To judge well, to comprehend well, to reason well, these are the essential activities of intelligence.”); Robert J. Sternberg, Beyond IQ: A Triarchic Theory of Human Intelligence (1985).
  \item See generally Howard Gardner, Frames of Mind: The Theory of Multiple Intelligences (1983) (describing multiple types of intelligence, including bodily-kinesthetic intelligence, interpersonal intelligence and musical intelligence); Howard Gardner, Intelligence Reframed: Multiple Intelligences for the 21st Century (1999) (same).
  \item See Binet & Simon, supra note 87, at 40 (describing the pedagogical method of measuring intelligence as judging intelligence “according to the sum of acquired knowledge”).
  \item Goldstein et al., supra note 82, at 108 (“Establishing a psychometric phenomenon of g does not mean that a psychological construct exists.” (internal citation omitted)).
  \item See Cianciolo & Sternberg, supra note 3, at 33–34 (“Binet’s interest in intelligence testing arose from the . . . practical concern of discriminating between people who could succeed academically and who could not . . . . Children shown via examination to have mental deficiencies were to be placed in special-education programs.”). Binet wrote an

\end{itemize}
psychologist who led the child taking the test through a series of tasks to assess the child’s “general potential” with a single score Binet called the child’s “mental age.” Binet’s test, particularly its approach of assigning a numerical score to an individual’s performance on a mental test, and Spearman’s theory of the “g” general intelligence factor had a strong influence on intelligence research and the design of mental tests for the rest of the twentieth century.

In 1916, at Stanford University, American psychologist Lewis Terman revised the “mental age” scale created by Binet and renamed it the “Stanford-Binet Intelligence Scales.” Terman’s Stanford-Binet intelligence test assigned test-takers an Intelligence Quotient (“IQ”) score—the ratio of their “mental” age and chronological age—based on their test performance. Terman’s new IQ test led to the widespread use of intelligence testing to sort individuals based on their test-predicted potential to perform particular jobs. During World War I, Terman, along with several other American psychometricians—Henry Goddard, Robert Yerkes, and Carl Brigham—successfully demonstrated that intelligence testing could be accomplished on a mass scale. The first mass-produced and mass-administered tests of intelligence—the Army Alpha and Beta exams—

---

93 Binet’s test brought “together a large series of short tasks, related to everyday problems of life (counting coins, or assessing which face is ‘prettier,’ for example), but supposedly involving such basic processes of reasoning as direction (ordering), comprehension, invention and censure (correction).” Stephen Jay Gould, The Mismeasure of Man 149 (1981) (internal quotation marks omitted).  
94 See e.g., Robert Sternberg, Handbook on Intelligence 17.  
95 Cianciolo & Sternberg, supra note 3, at 34 (“The first major revision was completed by Lewis Terman . . . . He translated the test (to the Stanford-Binet Intelligence Scales) for American use . . . .”). An individual’s IQ is defined as that person’s mental age (determined based on performance on the ninety Stanford-Binet tasks) divided by his or her chronological age and multiplied by 100. Id.  
96 Terman’s test remains influential to this day. Other standardized tests are deemed valid and reliable to the extent that their results align with psychometric properties of the Stanford-Binet IQ test. Id.  
97 Terman also advocated using mental tests to determine whether an individual possessed sufficient intelligence to perform particular jobs. Lewis M. Terman, The Measurement of Intelligence 17 (Ellwood P. Cubberley ed., 1916) (“Researches of this kind will ultimately determine the minimum ‘intelligence quotient’ necessary for success in each leading occupation.”).  
98 See, e.g., Cianciolo & Sternberg, supra note 3, at 35 (“The shift from one-on-one testing to group testing required substantial changes to intelligence test design. These changes included (1) the presentation of brief, written items in lieu of more complicated tasks requiring detailed instructions; (2) the replacement of examiner judgment with objective, right/wrong scoring techniques; (3) the imposition of time limits for test completion; and (4) the development of test problems appropriate for adults.”)
were used to test 1.75 million men.\footnote{Gould, supra note 93, at 195 (“Yerkes now had uniform data on 1.75 million men, and he had devised, in the Alpha and Beta exams, the first mass-produced written tests of intelligence.”).} Although the Army did not choose to rely on the results of the Alpha and Beta exams tests to sort individuals as rigidly as the psychologists who created the tests had advocated,\footnote{See id. at 194 (“I do not think that the army ever made much use of the tests.”).} the large data set produced by the Army exams laid the scientific foundation for a new college admissions test that would become integral to elite higher education admissions in the United States.\footnote{See id. at 195 (“[T]he major impact of Yerkes's tests did not fall upon the army . . . . Binet’s purpose could now be circumvented because a technology had been developed for testing all pupils. Tests could now rank and stream everybody; the era of mass testing had begun.”).}

\subsection*{1. Early Use of the SAT}

Princeton psychologist Carl Brigham became an important figure in the history of college admissions testing because he used the Army Alpha and Beta exams as a template for a test of an individual’s “scholastic aptitude”—the now famous Scholastic Aptitude Test (“SAT”). Brigham’s invention in the 1920s of a mass-market college admissions test contributed significantly to the rise of a “meritocratic ideology” in elite higher education admissions that began in the early 1930s and continues today. In 1934, when Harvard University announced a scholarship program to admit the most “naturally brilliant” young men from across the country—Harvard’s National Scholars Program—the program’s goal was to identify students with “the brains” to excel at Harvard but lacking the financial means to pay Harvard’s high tuition.\footnote{Lemann, supra note 31, at 28.} On the advice of Henry Chauncey, an associate dean who soon became the first President of today's Educational Testing Service (“ETS”),\footnote{The ETS is currently the company that designs and sells the world’s most widely used standardized tests. In addition to the SAT, ETS designs, sells, and administers the Preliminary SAT (“PSAT”), Advanced Placement (“AP”) tests, and GRE. See Tests & Products, ETS, http://www.ets.org/tests_products (last visited May 11, 2011) (listing the tests and products offered by ETS).} Harvard President James Conant decided to identify Harvard National Scholars using the newly developed product of psychometric research—the SAT.

Conant’s goal was to select high school students with the intellectual ability to become Harvard National Scholars without regard to their socio-economic class status. Essentially Chauncey convinced
Conant that Brigham’s new standardized test was a scientific breakthrough that made this possible. Hence, within thirty years of the publication of Charles Spearman’s seminal article on the “g” general factor of intelligence, Harvard University had adopted the practice of relying on the g-based SAT to select students.104

Conant set in motion an approach to elite higher education admissions that viewed mental testing as a path to upward mobility and national leadership for (white male) students who were among America’s intellectual elite. 105 Subsequently, the College Entrance Examination Board (“the College Board”) adopted the SAT to replace the subject-matter essay exams aligned with the curriculum of east coast preparatory schools that had previously been administered to applicants to elite private colleges such as Harvard, Yale, and Princeton.106 Notably, Brigham’s views about racial, ethnic, and class differences in intelligence107 did not stop university leaders like Harvard President

104 See LEMANN, supra note 31, at 43 (“[T]here was the possibility of creating a true national aristocracy to govern America . . . . Moreover, science . . . now offered in mental testing a way of selecting the country’s deserving new leaders. The SAT, in other words, would finally make possible the creation of a natural aristocracy.”).

105 While it is likely that Conant shared the prevailing view of non-whites as less intelligent than whites, Conant’s focus was the issue of class. The Harvard National Scholars Program was limited to white male applicants. See KARABEL, supra note 27, at 177 (noting that although the National Scholars program opened the doors of Harvard to working class whites, “the program discouraged black applicants and discriminated against those Jews who did apply”); see also LEMANN, supra note 31, at 47 (noting that “the most obvious departures from the American democratic ideal during the 1940s—legal segregation in the South, informal segregation elsewhere, and the relegation of women to a secondary position in society—went unmentioned by Conant in his writings”).

106 See KARABEL, supra note 27, at 44 (“Among the most important [measures] were the decision in 1905–1906 to replace Harvard’s own exams with those of the College Entrance Examination Board . . . .”).

107 See, e.g., CARL C. BRIGHAM, A STUDY OF AMERICAN INTELLIGENCE (1923). Carl Brigham is noted for studying racial and ethnic differences in the Army test results and advocating restrictions on immigration and the regulation of reproduction based on his findings. See GOULD, supra note 93, at 224–31 (describing Brigham’s belief in innate intelligence and stating that “the army data had its most immediate and profound impact upon the great immigration debate, then a major political issue in America, and ultimately the greatest triumph of eugenics”); see also, e.g. STERNBERG, supra note 87, at 28 (describing these psychologist as believing “that Africans and their descendants differed in intelligence from White people for genetic reasons” and identifying the contemporary writings of Richard J. Herrnstein and Charles Murray and Arthur Jensen as supporting this view as well). Carl Brigham was one of many central figures in the history of intelligence research, including British psychologist Sir Francis Galton, who believed that there are significant differences in the intelligence levels—average levels of “g”—of racial, ethnic, and class groups and that these group differences were primarily hereditary. Id. Today, contemporary psychometricians such as Arthur Jensen and Linda Gottfredson still argue in support of the “hereditability” of the “g” factor. See generally, ARTHUR R. JENSEN, THE G FACTOR: THE SCIENCE OF MENTAL ABILITY (Seymour W. Itzkoff ed., 1998) (concluding that intelligence is best understood as driven by the general intelligence factor); Linda S. Gottfredson, The
June 2011]  INTELLIGENT DESIGN  1261

James Conant and, later, the College Board from utilizing the SAT.\(^\text{108}\)

Himself an intellectual success story not born into the upper class, Conant sought to set in motion a new approach to elite higher education admissions that created a path to upward mobility and national leadership for intelligent men of middle class means who performed well on factorist tests like the SAT.

2. The Promise of the SAT’s Predictive Power

When first administered in 1926, the SAT offered “something entirely new: an easily scored, multiple-choice instrument for measur-
ing students’ general ability or aptitude for learning.”\textsuperscript{109} Brigham, the SAT’s creator, had conducted studies he claimed demonstrated that the SAT could predict a test-taker’s future academic performance.\textsuperscript{110} Based on its capacity to predict individual scholastic potential, the SAT college admissions test became a central feature of the “meritocracy” ethos now at the heart of American selective admissions. In historical context, the tenets of this new meritocratic ideal were a significant change from the class-restrictive status quo in elite higher education admissions.

A key principle (without regard to the exclusion of non-whites and white women) was that individuals with high general intellectual ability, not just members of the upper socioeconomic class, should be admitted to the nation’s top colleges if they possessed scientifically discernable superior mental ability.\textsuperscript{111} Prior to the creation of the SAT, middle and working class students had virtually no access to the nation’s elite private universities because only students who attended elite private secondary schools took courses in high school that prepared them to take the curriculum-specific essay exams administered in Latin by the College Board.\textsuperscript{112} Mental tests like Brigham’s SAT were touted as a new technology that could expand access to elite higher education beyond the upper classes.

The capacity of the SAT to predict future academic performance was central to Conant’s class-focused (but race- and gender-exclusionary) vision of allowing the academically talented to gain access to elite universities irrespective of their financial and social status. Analysis of the college grades of students selected as Harvard National Scholars based on their high SAT scores supported the theory that SAT scores could identify students who would excel in college

\textsuperscript{109} Atkinson & Geiser, supra note 4, at 666 (citing LEMANN, supra note 31).

\textsuperscript{110} LEMANN, supra note 31, at 33 (“Brigham had data to support the idea that the SAT could predict academic performance.”).

\textsuperscript{111} For instance, the presumption is that a public high school student has a better chance of performing well on the SAT, because it measures general intellectual ability, than he would on curriculum specific essay exams emphasizing subject-matter available to upper class students. \textit{See generally} KARABEL, supra note 27, at 2 (describing the history of selective admissions at Harvard, Yale, and Princeton and explaining that “academic merit” has only recently become a major factor in determining which students to select and that, dating back to earliest uses of the SAT as an admissions criteria, elite universities adopted admissions policies that afforded sufficient “discretion and opacity” to permit officials to “accept—and to reject—whomever they desired”).

\textsuperscript{112} \textit{See, e.g.}, \textit{id.} at 22–23 (“Both Yale and Princeton required that candidates pass examinations in both Greek and Latin, thereby effectively excluding most high school graduates, for only a handful of public schools offered both languages.”).
Many of Harvard’s SAT-identified National Scholars did go on to graduate cum laude, magna cum laude, and summa cum laude and to be elected Phi Beta Kappa. Harvard’s program boosted an “emerging national movement for the ‘scientific’ evaluation of academic potential” that soon led to the SAT being required of all applicants to leading colleges.

Thus, Harvard was among the first educational institutions to consider SAT scores based on their potential to predict future grades in college. Conant’s objective for the Harvard National Scholars Program can be fairly described as seeking to identify intellectual “diamonds in the rough,” individuals who possess the greatest intellectual ability and who exert the greatest amount of effort to apply that ability but lack the privileges of the aristocratic class. The underlying assumption is that these chosen meritocrats will go on to utilize their talents for the greater good and betterment of society.

First introduced at Harvard by Conant, the meritocracy-driven ideal of selective higher education admissions is now deeply entrenched in American society. Today’s most selective colleges and universities essentially ascribe to a theory very similar to the one that prompted Harvard to first use SAT scores as admissions criterion—the theory that individuals’ scores on tests like the SAT should be

113 KARABEL, supra note 27, at 140 (“With each additional group of National Scholars who compiled outstanding records, it seemed increasingly clear that it was possible, with the assistance of the new science of psychometrics, to predict which applicants would prove to be brilliant scholars.”); LEMANN, supra note 31, at 39 (stating that eight of the ten Harvard National Scholars for the class of 1938 were elected to Phi Beta Kappa).

114 KARABEL, supra note 27, at 177 (“Of the 10 recipients who graduated in 1938, 5 graduated summa cum laude, 3 magna cum laude, 2 cum laude, and 8 were elected to Phi Beta Kappa.”). But see id. (explaining that Harvard scholarship students were required to maintain a high GPA in order to keep their scholarship and remain in college). The academic success of the National Scholars might also be attributed to the fact that, during the 1930s and 1940s, college academics were far less competitive at elite private schools like Harvard than is the case today. See id. at 21 (noting that Harvard’s student culture was "largely hostile to academic exertion"). Except for less affluent students attending such schools on scholarships, most “Harvard men” were competing to achieve high positions in non-academic, extracurricular clubs and teams. See id. at 17 (“[T]he academic side of the college experience ranked a distant third behind club life and campus activities.”).

115 See id. at 198, 266, 425 (describing the pressure on Harvard, Yale, and Princeton to admit students based on SAT scores). The rationale that drove other elite institutions to adopt the SAT as an admissions criterion still resonates today—the SAT is a more uniform, thus more fair, basis for comparing students from various parts of the country who attend schools of varying quality than high school grades or assessments of students’ mastery of specific subject matter that may not have been taught at their high school.

116 Id. at 198. On December 14, 1941, Harvard and other elite private colleges established a policy requiring the SAT for all applicants, not just scholarship candidates. Id. at 178.
considered as an admissions criterion because of the capacity of the SAT to predict test-takers’ future grades. Psychometricians now regularly produce institution-specific validity studies—studies that compare the entering SAT scores of admitted students—and institution-specific regression equations that reflect the statistical relationship between SAT scores and first-year grades.\textsuperscript{117} The “correlation coefficient,” \( r \), between the SAT and college performance is derived by comparing an accepted student’s SAT score with his or her overall GPA at the end of the freshman year of college. The square of the correlation coefficient, the “coefficient of determination” \( r^2 \), is a measure of the strength of the SAT’s capacity to predict that a higher-scoring test-taker will have higher grades in college than a lower-scoring test-taker.

B. Utility and Racial Problems of Factorist Tests: The SAT Example

Selective public and private universities still operate according to the principle that led Conant to adopt the SAT as an admissions criterion for Harvard National Scholars in the 1930s—that the great benefit of factorist admissions tests like the SAT, GRE, LSAT, and MCAT is the incremental improvement in prediction they contribute to admissions decision-making as compared to relying on non-test criteria exclusively.\textsuperscript{118} As further explained in this section which again uses the SAT as an example, inclusion of the SAT as an admissions criterion increases the percentage of “variation” in test-takers’ future college grades explained (or predicted) from approximately 18% relying on high school grades alone to roughly 23% if both SAT scores and high school grades are considered. During the century that factorist tests have been in existence, it has been this incremental im-

\textsuperscript{117} How institutions use such equations varies dramatically. If they so choose, institutions can use their institution-specific regression equation to develop an equation for combining an applicant’s SAT score with his or her high school GPA—a numerical composite index score which can be used to compare and rank applicants quantitatively. Although the high school GPA of applicants is more predictive of college success (defined by first-year grades) than SAT scores alone, many studies have found that relying on a combination of SAT and high school GPA is more predictive than using high school GPA alone. See, e.g., REBECCA ZWICK, FAIR GAME?: THE USE OF STANDARDIZED ADMISSIONS TESTS IN HIGHER EDUCATION 79–90 (2002).

\textsuperscript{118} The GRE, LSAT, and MCAT play a similar role in graduate, law, and medical school admissions. See, e.g., LANI GUNIER, MICHELLE FINE & JANE BALIN, BECOMING GENTLEMEN: WOMEN, LAW SCHOOL, AND INSTITUTIONAL CHANGE 40 (1997) (showing the relationship between LSAT scores and first-year law student GPAs); 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, 5 (1997) (examining the role the LSAT and grades play in law school admissions).
provement in predictive power that has justified their continued use in selective admissions.

However, another long-standing characteristic has plagued traditional factorist tests since their creation—the consistent and statistically significant differences in the averaged factorist test scores of certain racial groups.119 Lack of certainty as to the explanations for the racial differences in factorist test scores combined with the controversial tenet ascribed to by many g-oriented intelligence theorists that such differences are hereditary and virtually immutable120 have led some to question the fairness of factorist tests as admissions criteria. The fact that the averaged test scores of African-American and Latino students is lower than the averaged scores of Asian-American and white students often lead to accusations that considering factorist test scores in selective admissions unfairly excludes otherwise qualified African Americans and Latinos. Conversely, rejected whites point to the fact that the averaged scores of some admitted African Americans and Latinos are lower than the averaged scores of white students to claim that selective universities unfairly rely less on factorist test scores in assessing such applicants.121

1. Modern Predictive Power of the SAT

Today, the major appeal of mental testing, particularly selective higher education admissions testing, is that it confers a degree of scientific credibility to the process of comparing individuals’ intellectual abilities that is lacking when admissions decisions are based solely on grades and other non-test criteria. Though long criticized for doing little more than measuring differences in test-takers’ social and economic opportunities, a major reason that standardized tests like

---

119 Whites as a racial group have, on average, higher scores on conventional factorist admissions tests than certain non-white racial groups including African-Americans, Latinos and some subcategories of the Asian American racial group. Although it is not true of all categories of Asian Americans, some Asian racial groups score higher than whites on some standardized admissions tests. CiancioLO & STERNBERG, supra note 3, at 125–26 (discussing “substantial Asian superiority in academic achievement tests (e.g. grades and achievement tests, such as the SAT”).

120 See, e.g., richARD J. HERNEsTEn & charles murray, the bell curve: intelligence and class structure in American life (1994) (arguing that American society is divided into a cognitive elite and cognitive underclass based on IQ scores resulting from hereditable genetic disadvantages that cannot be overcome by public policy interventions). But see chrisToPhEr jEncks & mEriDIdH phillips, the black-white test score gap 6–12 (1998) (compiling various studies concluding the black-white test score gap is not genetic in origin but instead the result of the cultural, economic, and educational differences among racial groups).

121 See infra note 216 and accompanying text.
the SAT, GRE, LSAT, and MCAT have maintained their prominence as admission criteria is the fact that psychometricians—measurement psychologists—have consistently produced statistical evidence that such tests are scientific predictors of test-takers’ future performance at the educational institutions to which they apply. The basis upon which standardized tests offer this scientific support of admissions decisions is the statistical correlation that psychometricians have found between test-takers’ scores on admissions tests and their actual grades in college, graduate, and professional schools.

To a large extent, it is the statistical correlation between test-takers’ scores on admissions tests and test-takers’ first-year grades—the “predictive power” of admissions tests—that forms the scientific justification for treating applicants’ test scores as useful proxies for admissions-related merit. Studies of the correlation between test-takers’ SAT scores and early college performance consistently have been found to be statistically significant. The statistical principles that inform test score use are the basis for the conclusion that there is a relationship between SAT scores and future grades that cannot be explained by chance alone. Yet, although college grades and SAT scores are correlated, that correlation is far from perfect and not as

---

122 See, e.g., Linda S. Gottfredson & James Crouse, Validity Versus Utility of Mental Tests: Example of the SAT, 29 J. VOCATIONAL. BEHAV. 363, 365–66 (1986) (describing Crouse’s longitudinal studies’ conclusions that aptitude tests predict later academic performance). More precisely, the practical utility of college and graduate school admissions tests is the extent to which scores on such tests are “correlated” to test-takers’ first-year grades at a particular educational institution. See discussion supra Part I.

123 It is worth noting that the most commonly used “outcome criterion”—test-takers’ first-year grades—is regularly criticized as having little practical significance. See, e.g., West-Faulcon, supra note 5, at 1116 (“The Educational Testing Service . . . reports that a high school student’s SAT score explains approximately thirteen percent of the variance in first-year college grades, less than would be explained if universities relied on high-school grades alone.”). In fact, researchers have gone to great lengths to identify more “construct valid” and practically significant outcomes. See generally, Marjorie M. Shultz & Sheldon Zedeck, Predicting Lawyer Effectiveness: A New Assessment for Use in Law School Admissions Decisions 77, 80–81 (July 31, 2009), CELS 2009 4th Annual Conference on Empirical Legal Studies Paper (describing the development of an assessment test designed to predict post-law school “lawyer effectiveness” instead of first-year grades, the criterion predicted by the LSAT).

124 See, e.g., Zwick, supra note 117, at 85–86 (finding a correlation of .36 when using only the SAT as a predictor). SAT scores can be fairly said to correlate to test-takers’ first-year GPA and also, but to a lesser degree, test-takers’ later college performance (GPA after the freshman year). Id.

125 If the SAT were a perfect predictor of test-takers’ future grades in college, universities would be able to predict an individual’s grades simply by knowing his or her SAT score. See, e.g., Michael Selmi, Testing for Equality: Merit, Efficiency, and the Affirmative Action Debate, 42 UCLA L. REV. 1251, 1263 (1995) (making similar observation about the relationship between employment test scores and employees’ performance ratings). Whenever a
good as the correlation between high school grades and college grades. 126

The SAT’s “practical predictive power”—the percentage of variation in test-takers’ first-year grades that is explained by test-takers’ SAT scores—is the “squared correlation coefficient” ($r^2$). Because SAT scores explain approximately 13% of the variance in first-year college grades, 127 as much as 87% of the variation in scholastic success during the first year of college is not explained by an individual’s SAT score. 128 In other words, the fact that the SAT college admissions test,

correlation between two variables is less than perfect, using one of those two variables to predict the other will result in errors. Id. at n.42 (describing false negatives as misclassifications). Statistical theory uses the terms “false positive” and “false negative” to describe the instances where the presumed relationship between the variables is not present—in the context of mental testing, when rank-order differences in test scores do not translate into better relative academic performance. Id. An applicant is a “false positive” if his or her actual performance falls short of what is normally expected of someone with a high SAT score. Id. In the reverse, a “false negative” is an applicant whose SAT score is too low to qualify him or her for admission but who, if admitted, would succeed academically at a level that far exceeds SAT-based expectations. Id. In comparing the predictive accuracy or predictive power of two tests that are otherwise equally reliable, the test that results in the fewest false positives and fewest false negatives is a better predictive tool. Id.

Admissions tests explain even smaller percentages of the variation in applicants’ later college grades (e.g. GPA at graduation). Atkinson & Geiser, supra note 4, at 672 (“[T]here is little difference among the major national tests in their ability to predict student performance in college.”).

126 See Zwick, supra note 117, at 116 tbl.5-2. Because of the large number and wide variation in studies of the SAT’s predictive effectiveness, and the fact that testing experts disagree as to whether correlation coefficients should be “corrected” to compensate for “restriction of range,” see Selmi supra note 125, at 1266–67, reported correlations between SAT scores and grades vary from 0.36 to as high as 0.65. See Zwick, supra note 117, at 84–86. Zwick reports the results of an ETS report that finds an overall correlation of SAT verbal and math with college GPA of 0.36 and an overall correlation of high school GPA with college GPA of 0.39. Id. at 116. The percentage of the variation ($r^2$) in college GPA explained is calculated by squaring the correlation coefficient ($r$) and multiplying by 100. Id.

127 In that same study, test-takers’ overall high school GPA by itself explained roughly 16%—more than SAT score—but still left 84% of variation in early college grades unexplained. The fact that students’ high school GPA is consistently a better predictor of both their freshman and four-year college grades has led to repeated challenges of the incremental predictive value of the SAT as an admissions criterion. See, e.g., Saul Geiser & Maria Veronica Santelices, Validity of High School Grades in Predicting Student Success Beyond the Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes, CTR. FOR STUDIES IN HIGHER EDUC. 1, 1 (2007), http://csh.berkeley.edu/publications/docs/ROPS.GEISER_SAT_6.13.07.pdf (“High-school grades are often viewed as an unreliable criterion for college admissions, owing to differences in grading standards across high schools, while standardized tests are seen as methodologically rigorous, providing a more uniform and valid yardstick for assessing student ability and achievement. The present study challenges that conventional view.”); Susan Sturm & Lani Guinier, The Future of Affirmative Action: Reclaiming the Innovative Ideal, 84 CALIF. L. REV. 953, 974 (1996) (“It is widely recognized that high school grades are more predictive of
considered alone, explains approximately 13% of the variation in a test-taker’s college grades means that, of the factors that contribute to which freshmen have higher or lower GPAs at the end of their first year, the SAT explains 13% and the other 87% of what explains why applicants perform better academically relative to one another is left unaccounted for and unexplained by consideration of applicants’ SAT score by itself.129

When, as is typically the practice of universities, both high school GPA and SAT scores are considered together, more of the variation in the differences in freshman college grades is explained. Even when a factorist test like the SAT is used in conjunction with high school grades to predict college performance, over 70% of the variance is unaccounted for and unexplained. Universities have generally made the policy judgment that having some scientific basis for college freshman-year grades than the SAT. Perhaps even more significant is the extremely small increase in predictiveness gained by using the SAT in conjunction with high school grades.

Critics also point to studies that consistently find the correlation between white test-takers’ scores to be higher than the correlations for non-whites as proof that the SAT predicts less well for non-whites. Nevertheless, most psychologists are of the view that the predictive effectiveness of the SAT is similar for different racial groups. See, e.g., The Rainbow Project, supra note 21, at 322 (“[T]he predictive effectiveness of the SAT-C[omined Score] varies from the highest (.64) for White students to the lowest (.50) for Native American students, with Asian American (.65), Black (.62), and Hispanic (.53) students taking intermediate positions in the order specified here.”). A major reason why selective institutions rely on the SAT as opposed to relying exclusively on applicants’ prior grades is a concern over the lack of scientific reliability of high school grades: teachers’ grading criteria are believed to be highly variable and inconsistent. Thus, it is often argued that the SAT and other standardized admissions tests serve as a critical check on grades without which grades might eventually become useless as admissions criteria; the fact that students take national, standardized admissions tests discourages teachers from unjustifiably inflating students’ grades to increase their college admissions prospects. See, e.g., West-Faulcon, supra note 5, at 114 (“The SAT remains such a prominent criterion in college admissions because of its unique capacity to provide a standardized national yardstick for comparing students from high schools across the country . . . .”); see also ROBERT K. FULLINWIDER & JUDITH LICHTENBERG, LEVELING THE PLAYING FIELD: JUSTICE, POLITICS, AND COLLEGE ADMISSIONS 119 (2004) (“What a standardized test such as the SAT I or the ACT provides is an objective measure based on a comprehensive national cross-section of college-goers.”).

I d. at 85–86.

129 See, e.g., Zwick, supra note 117, at 85 (calculating standardized tests’ predictive effectiveness). Zwick states:

The regression analysis using only high school GPA as a predictor yielded a moderately high correlation of .39 [fifteen percent of the variation in college grades]; using only the SAT produced a correlation of .56. When high school GPA, SAT math, and SAT verbal scores were used in combination, the correlation rose to .48 [twenty three percent of the variation in college grades], yielding an ‘SAT increment’ of .09 (.48 minus .39). These findings parallel the results of many other test validity analyses in two basic ways. First, prior grades alone were slightly more effective in predicting subsequent grades than were admissions test scores alone. Second, adding test scores to prior grades improved the prediction. Id. at 85–86.
admissions decisions is better than relying solely on non-test criteria. A major rationale for relying on standardized admissions tests in addition to prior grades is their incremental predictive power when considered in combination with applicants’ grades. In the context of selective college admissions, taking SAT scores into account in addition to high school GPA adds an increment of about six to eight percentage points to the percentage points of variation in applicants’ first-year college grades explained by high school grades considered alone.\textsuperscript{130}

Thus, at the end of the roughly hundred-year period that mass-marketed standardized tests have been in existence, their predictive power still leaves substantial room for improvement.\textsuperscript{132} College and graduate school admissions tests measure only a relatively small portion of the differences in test-takers’ future academic success and, thus, possess far less than perfect (100\%) predictive power; their predictive power still leaves substantial room for improvement because they leave more of the variation in intelligence and future academic success unexplained than they actually explain. Accordingly, the strength of prediction possible using SAT scores along with grades suggests a need for further innovation in testing.\textsuperscript{133}

\textsuperscript{130} Using Zwick’s data, my calculation of 23\% of the variation in college grades as explained when SAT scores are considered in combination with high school grades (based on the correlation (\(r\)) of .48 and coefficient of determination (\(r^2\)) of .23) is 8\% higher than the 15\% of the variation in college grades explained by SAT scores when considered alone (based on the correlation (\(r\)) of .39 and coefficient of determination (\(r^2\)) of .15). See, e.g., Zwick, supra note 117, at 85–86 (“The regression analysis using only high school GPA as a predictor yielded a moderately high correlation of .39; using only the SAT produced a correlation of .36. When high school GPA, SAT math, and SAT verbal scores were used in combination, the correlation rose to .48 . . . .”); Atkinson & Geiser, supra note 4, at 666 (“In our studies at the University of California, for example, we have found that admissions tests add an increment of about 6 percentage points to the explained variance in cumulative college GPA . . . .”).

\textsuperscript{131} Although it is not the focus of this Article, the fact that such a small percentage of incremental prediction is gained by relying on admissions tests could be the basis of a policy argument that universities should not rely on them or a legal argument that such reliance is not an “educational necessity.” See West-Faulcon, supra note 5, at 1126–28 (“[i]n our studies at the University of California, for example, we have found that admissions tests add an increment of about 6 percentage points to the explained variance in cumulative college GPA . . . .”).

\textsuperscript{132} Atkinson & Geiser, supra note 4, at 672 (stating that “[a]fter decades of predictive validity studies, our best models still account for only about 25\% to 30\% of the variance in outcome measures such as college GPA”).

\textsuperscript{133} See Atkinson & Geiser, supra note 4, at 672 (“Our ability to predict student performance in college on the basis of factors known at point of admission remains relatively limited.”).
2. Differences in Averaged SAT Scores of Racial Groups

When the SAT scores of members of racial groups are averaged, numerous studies have found significant differences in those averages. While their added contribution to the prediction of applicants’ future academic performance is essentially the greatest benefit offered by factorist tests as an admissions criterion, the fact that racial groups, on average, perform differently on such tests has long prompted questions as to whether their use is of sufficient net substantive value to justify their impact on lower-scoring racial groups. The impact of the racial gap in scores on factorist tests like the SAT is exacerbated by the fact that the size of the gap is significant and is characteristic of virtually all factorist tests.

Research studies have found the averaged score of African-American test-takers on the SAT is about one full standard deviation lower than the averaged score of white test-takers. The meaning of measured racial differences in test scores should take into account the fact that "that there is disagreement as to whether race is a biologically meaningful concept." CIANCIOLO & STERENBERG, supra note 3, at 121. "An alternative to the biological approach is to define race as a social construct rooted in historical and anthropological context." James L. Outtz & Daniel A. Newman, A Theory of Adverse Impact, in ADVERSE IMPACT: IMPLICATIONS FOR ORGANIZATIONAL STAFFING AND HIGH STAKES SELECTION 57–58 (James Outzz ed., 2010) (describing authors’ agreement with “the premise that race is a social construct” and their conceptualization of race “to include group-level shared perceptions/meanings, resulting from common societal experiences, as well as individual-level constructs (e.g. unique personal meanings drawn from the common experience”).

See The Rainbow Project, supra note 21, at 322 (“[I]n comparison with White Students on average, African American students scored about one full SD [standard deviation] lower . . . .”). Studies have found similar differences in LSAT scores. See, e.g., LINDA F. WIGHTMAN & DAVID G. MULLER, LAW SCH. ADMISSION COUNCIL, AN ANALYSIS OF DIFFERENTIAL VALIDITY AND DIFFERENTIAL PREDICTION FOR BLACK, MEXICAN-AMERICAN, HISPANIC, AND WHITE LAW SCHOOL STUDENTS 9 tbl.6a (1990) (comparing the means and standard deviations of LSAT scores and GPAs for whites and African Americans); Lempert, et. al., supra note 11, at 244 (reviewing Steven T. Ziliak and Deirdre N. McCloskey’s book on statistical significance which provides data on racial differences on the LSAT); William C. Kidder, The Struggle for Access from Sweatt to Grutter: A History of African American, Latino, and American Indian Law School Admissions, 1950–2000, 19 HARV. BLACKLETTER L.J. 1 (2003) (discussing generally the history of law school admissions and the LSAT as it applies to different racial groups); Linda F. Wightman, The Consequences of Race-Blindness: Revisiting Prediction Models with Current Law School Data, 53 J. LEGAL EDUC. 229, 244–45 (2003) (discussing the difference in LSAT scores between white and African American applicants); Wightman, supra note 118, at 36 (“The black students in this sample came to law school with UGPAs that are, on average, one standard deviation below those of white students and LSAT scores that average more than one-and-a-half standard deviations below.”); see also JENCKS & PHILLIPS, supra note 120, at 1–51 (discussing generally the gap between black and white standardized test scores from kindergarten through adulthood); William D. Henderson, The LSAT, Law School Exams, and Meritocracy: The Surprising and Undertheorized Role of Test-Taking Speed, 82 TEX. L. REV. 975, 978 (2004) (“The LSAT also
Latino SAT score is slightly lower than one standard deviation (0.9) less than the white averaged score. The Native-American averaged SAT score is about half (0.5) a standard deviation lower than the averaged white score. In contrast, the averaged score of Asian students (not disaggregating particular Asian subgroups) is slightly higher, by 0.2 standard deviation, than the white average—averaged Asian SAT scores are higher than the white averaged scores on SAT math tests but lower on SAT verbal and writing tests. Multiple explanations for the racial differences in factorist test scores have been suggested: cultural, economic, and hereditary differences among racial groups, the psychological impact of racial stereotypes about intellectual differences among racial groups, and racially disparate impacts of test development procedures and item selection processes have all been identified as factors that may contribute to “the racial gap” in factorist test scores. Significantly, the next Part describes recent research designed to define and measure intelligence more completely than $g$-based theories that has spawned tests with smaller racial gaps in scores and greater predictive power than the

136 See, e.g., The Rainbow Project, supra note 21, at 322 (“When all SAT scores were aggregated . . . in comparison with White students . . . , Latino students scored 0.9 SD [standard deviation] lower . . . .”) . Studies have shown similar results with LSAT scores. See, e.g., Wightman, The Consequences of Race-Blindness, supra note 135, at 245 tbl.8 (listing the differences in standard deviations of LSAT scores of Hispanic and white law school applicants).

137 See, e.g., The Rainbow Project, supra note 21, at 322 (“Native Americans scored about half a SD [standard deviation] lower [than white students].”) . For similar differences in LSAT scores, see, for example, Wightman, The Consequences of Race-Blindness, supra note 135, at 245 tbl.8 (showing the differences in standard deviations of LSAT scores between white and Native American law school applicants).

138 The Rainbow Project, supra note 21, at 322 (“[Asian students] scored higher than White students by about .03 (SAT I) to .07 (SAT II) SDs [standard deviations] on the math tests, but about a third (SAT I) to half a (SAT II) SD [standard deviation] lower on the verbal/writing tests.”).

139 See, e.g., supra note 120.

140 See Claude M. Steele, A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance, 52 AM. PSYCHOL. 613, 613 (1997) (describing research studies finding that societal stereotypes about groups can influence the intellectual functioning and identity development of individual group members); see also Claude M. Steele, WHISTLING VIVALDI: AND OTHER CLUES TO HOW STEREOTYPES AFFECT US 4 (2010) (“[B]y imposing on us certain conditions of life, our social identities can strongly affect things as important as our performances in the classroom and on standardized tests . . . .”).

141 William C. Kidder & Jay Rosner, How the SAT Creates “Built-In Headwinds”: An Educational and Legal Analysis of Disparate Impact, 43 SANTA CLARA L. REV. 131, 122–134 (2002) (describing how the manner in which SAT questions are developed and selected contributes to the significant disparate impact the SAT has on African-American and Latino test-takers).
more racially-skewed conventional factorist tests—an improvement in the technology of mental testing.

III. NEW UNDERSTANDINGS OF INTELLIGENCE: ABOVE AND BEYOND "g"

The theory of successful intelligence provides one basis for improving prediction and possibly for establishing greater group equity. It suggests that broadening the range of skills tested to go beyond the analytical and memory skills typically tapped by the SAT, to include practical and creative skills as well, might significantly enhance the prediction of college performance beyond current levels.

. . . .

The SAT is based on a conventional psychometric notion of cognitive skills . . . . But perhaps the time has come to move beyond conventional theories of cognitive skills . . . . [T]he triarchic measures alone approximately double the predicted amount of variance in college GPA when compared with the SAT alone . . . .

—Robert J. Sternberg

Modern theories of intelligence can be very roughly grouped into two categories—modern g-oriented theories of intelligence and modern theories of intelligence rejecting (or, at least, not dependent upon) the existence of "g." A broad range of non-factorist theories of intelligence conceive of intelligence as comprised of something more than or other than "g." These non-g-based theories, such as the CHC theory of cognitive abilities, Sternberg’s triarchic theory, the PASS Theory, and Multiple Intelligence Theory purport to define intelligence more completely than the conventional g-based theory.

142 The Rainbow Project, supra note 21, at 322, 344 (citation omitted).

143 In the later twentieth century, Charles Spearman’s original g-based theory of intelligence has been updated. These modern factorist theories of intelligence still place the "g" general factor at the top of tiered structures of mental abilities as the most significant factor. Two of the most prominent modern hierarchical psychometric theories of intelligence are Vernon and Cattell’s theory, introduced in 1971, that divides "g" into "fluid" and "crystallized" intelligence and Carroll’s "three stratum" theory, set forth in 1993, that places "g" general intelligence in "Stratum III"—at the top of a three-tier hierarchy of mental abilities. CIANCIOLO & STERNBERG, supra note 3, at 6–8.

144 Goldstein et al., supra note 82, at 102–09 (describing such theories). However, the debate over the correctness of multi-dimensional versus g-centered theories of intelligence is ongoing. For critiques of Sternberg’s and other alternative theories, see Linda S. Gottfredson, Practical Intelligence in Everyday Life, 29 INTELLIGENCE 363, 363–65 (2001) (book review) (arguing that the research approach employed was flawed); Arthur R. Jensen, Test Validity: g versus Tacit Knowledge, 2 CURRENT DIRECTIONS IN PSYCHOL. SCI. 9, 9 (1993)
Modern research on intelligence has resulted in the development of tests that do a better job than traditional g-based factorist tests of predicting test-takers’ future performance.\(^{145}\) Plus, that greater predictive power comes with an added bonus—smaller racial group differences in scores on such tests.\(^{146}\) The increasingly large body of intelligence research described in this Part increases selective universities’ vulnerability to the type of Title VI test deficiency administrative complaints introduced in Part I of this Article\(^{147}\) and examined further below.

A. Critiques of “g”

The conception of intelligence as “multi-dimensional” predates Spearman’s theory of “g.”\(^{148}\) Influential psychologists who disagreed with Spearman’s conception of intelligence as “g” included American psychologist Louis L. Thurstone who conceived intelligence as comprised of seven distinct but interrelated factors\(^{149}\) and Joy Paul Guilford who, rejecting “g,” proposed that intelligence involved over one hundred different factors.\(^{150}\) Presently, some modern intelligence experts assert that g-oriented factorist theories of intelligence have been seriously flawed from their inception.\(^{151}\)

A central critique of the factorist approach to describing the nature of intelligence is that it offers no empirical—only a mathemati-
An additional flaw of the mathematical process of factor analysis according to its critics is the fact that different psychometricians have used it to reach drastically different conclusions as to the number of factors of intelligence—some factorists have concluded that the human mind has as few as two factors while other factorists have used the same method and found as many as 120 mental factors. Today, factor analysis can be contrasted with more empirically-oriented cognitive psychology approaches to intelligence theory. Unlike factorist theories of intelligence, cognitive theories conceive intelligence as “cognitive processes” and seek to observe and measure intelligent behavior in the real world.

Beyond its mathematical focus, the g-oriented factorist theory of intelligence is also subject to particular criticism because it ultimately defines “g” as what g-oriented factorist tests measure and because the only criterion for a test to be deemed a measure of “g” under the psychometric perspective is that its score results align with the preconceived g-oriented hypothesis about the nature of intelligence. The essentially trial and error experimentation by which the components of early g-loaded tests were selected—retaining test questions and arrays when their results correlated to other g-loaded tests—is
demonstrative of the circuitous logic and weak empirical and theoretical underpinnings of the psychometric perspective that have led to modern efforts to replace or augment the theory of “g.”

B. Examples of Innovation in Intelligence Theory and Testing

Strikingly, the research of modern intelligence theorists and testing experts has revealed that conventional g-based standardized tests are not as predictive of test-takers’ future performance as tests designed according to non-g-based theories. Seeking to test the hypothesis that at least some of the racial differences in factorist tests scores do not reflect actual racial differences in test-takers’ intelligence and future academic success, these researchers have undertaken “efforts to develop valid tests of intelligence with reduced racial differences.”159 The thinking behind such research is that, in contrast to conventional g-oriented testing, “tests developed based on [more] sound theory could result in reducing deficiency- and contamination-related factors that might contribute to racial differences observed [in test scores].”160 By designing “more theoretically based tests such as those that focus on measuring key factors of intelligence (e.g., fluid reasoning, general memory and learning),” modern intelligence theorists believe it is possible to create tests that both do a better job of measuring intelligence and that demonstrate “lower racial differences” in scores.161

For instance, when revised “to better fit” the dimensions of the intelligence theory of psychologist John Carroll,162 traditional intelligence tests such as the Woodcock-Johnson III, Stanford-Binet 5, and WISC-IV tests had significantly decreased racial differences in scores without decreasing these tests’ capacity to measure intelligence.163 Instead of the one standard deviation typically reported as the differ-

160 Goldstein, supra note 82, at 125.
161 Id. at 124.
162 Id.
163 Id.
ence between the scores of African-American and white test-takers, revisions to the tests that aligned with Carroll’s well-regarded intelligence theory substantially reduced the gap between African-American and white test scores. Instead of one standard deviation between the scores of the two racial groups, the difference was cut in almost half in some instances to 0.54.

Another example of innovation in mental testing is the Cognitive Assessment System (“CAS”) designed based on the PASS theory of intelligence. PASS is a cognitive model of intelligence focused on performance that delineates four main factors as “the cognitive building blocks of human intellectual functioning.” The CAS test is an intelligence test for children and adolescents that is predictive of test-takers’ achievement in school settings. Although similar to traditional intelligence tests, the CAS test shows “much lower racial differences than found with other traditional tests of intelligence.” In contrast to the often reported one standard deviation between African-American and white test-takers, the CAS test designed according to the PASS theory of intelligence has been reported to have a black-white score difference of only 0.26 standard deviation—only a little more than one-fourth the size of the racial gap typically reported for g-oriented intelligence tests.

Also notable is the fact that similar innovations are taking place in the development of employment tests. The Siena Reasoning Test (“SRT”), designed based on the intelligence theory of Joseph Fagan, has predictive power “equal to or better than” traditional g-based cognitive tests and consistently has significantly smaller racial differences in scores—“yielding, for example, black-white mean differences that ranged from approximately 0.00 [no racial difference whatsoever] to 0.40 standard deviation [less than half the traditional black-white difference on g-based tests]."

164 See supra Part II.C.
165 Goldstein, supra 82, at 124 ("In fact, Wasserman and Becker reported racial differences below the 1 SD typically reported for some mainstream tests (e.g., black-white SD differences . . . . ranged from 0.54 to 0.73).").
166 Id. at 125.
169 Goldstein, supra note 82, at 125.
170 Id. at 127.
171 Id.
Lastly and particularly relevant to the analysis in this Article, similar research has also been undertaken in the context of college admissions testing. Psychologist Robert Sternberg’s project with the College Board produced similar findings when he better fitted the SAT to a multi-dimensional theory of intelligence—his own triarchic theory. The new college admissions test had increased power to predict test-takers’ future performance and reduced racial differences as compared to the comparable g-based SAT college admissions test.\(^{172}\)

C. A More Intelligent College Admissions Test?

Whereas traditional factorist definitions of intelligence conceptualize intelligence as structured around the “g” general factor, Robert Sternberg’s theory defines “successful intelligence” as comprised of three quantifiable mental abilities or intelligences: analytical intelligence, practical intelligence, and creative intelligence.\(^{173}\) Measuring “analytic intelligence” is a quantification of the application of abilities to relatively “familiar problems” of an “abstract nature.”\(^{174}\) As defined under the triarchic theory, analytical intelligence is involved when skills are used to analyze, evaluate, judge or compare and contrast. “Practical intelligence” involves “skills used to implement, apply, or put into practice ideas in real-world contexts.”\(^{175}\) Sternberg describes practical intelligence as the utilization of “tacit knowledge”—knowledge that one is “not explicitly taught” and that is often “not even verbalized” but that is necessary to work “effectively” in an environment.\(^{176}\) Lastly, the theory asserts that “creative intelligence” is the

---

\(^{172}\) Id. at 126(describing Sternberg’s Rainbow Project and findings).

\(^{173}\) Id. at 126(describing Sternberg’s Rainbow Project and findings).

\(^{174}\) Id. at 325.

\(^{175}\) See Kristin Garrigan & Jonathan Plucker, New and Emerging Theories of Intelligence, HUMAN INTELLIGENCE, http://www.indiana.edu/~intell/emerging.shtml (last modified July 25, 2007) (describing Sternberg’s conceptions of intelligence). Instead, the three intelligences of triarchic theory are perpetually developing abilities. Id.

\(^{176}\) See Robert J. Sternberg, WICS: A Model of Positive Educational Leadership, in THE ESSENTIAL STERNBERG: ESSAYS ON INTELLIGENCE, PSYCHOLOGY, AND EDUCATION 394 (James C. Kaufman & Elena L. Grigorenko eds., 2009) (describing “work-related” problems used to measure “tacit knowledge”); see also id. at 395 (describing findings that tests of tacit knowledge typically show no correlation with IQ tests, but “predict performance on the job as
application of abilities to “novel” situations and problems. Creative intelligence is involved when “skills are used to create, invent, discover, imagine, suppose or hypothesize.”

Sternberg has created both a multiple-choice test of the three aspects of intelligence conceived under his triarchic theory, Sternberg’s Triarchic Abilities Test and several types of “performance task” tests to measure practical and creative intelligence. All of these “triarchic tests” seek to test intelligence “more broadly” in order to better predict test-takers’ future grades with less racial disparity in scores than traditional factorist tests. In short, Sternberg’s theory is that by measuring additional fundamental facets of intelligence—“creative” and “practical” intelligence—as well as the “analytic” aspect of intelligence measured by traditional g-based factorist tests like the SAT, the systems-based triarchic tests advance the technology of mental testing; they attempt to measure mental ability more broadly and completely than conventional standardized tests.

In a study called the “Rainbow Project,” Sternberg’s goal was to “construct-validate” the triarchic theory of successful intelligence and “also to show its usefulness in a practical prediction situation.” The study was successful in accomplishing both goals. In the 2006 volume of Intelligence, Sternberg reported that his contemporary systems theory of intelligence, as operationalized in the form of the Sternberg Triarchic Abilities Test and performance measures, did indeed prove to be “construct valid” and “useful in increasing predictive validity, and, at the same time, in reducing ethnic group differences in scores.”

---

177 Id. at 393 (“We presented 80 individuals with novel kinds of reasoning problems . . . .”).  
178 The Rainbow Project, supra note 21, at 325. Creative intelligence problems ask test-takers to write stories, create art, design advertisements and develop science products. See WICS: A Model of Positive Educational Leadership, supra note 176, at 414 (“Individuals were asked to create products in the realms of writing, art, advertising, and science.”). A typical creative intelligence writing question might require the test-taker to choose from two titles, such as “Beyond the Edge” or “The Octopus’s Sneakers,” and to write a short story based on the title they choose. An art-related creative intelligence question might ask test-takers to produce an art composition with either the title “The Beginning of Time” or “Earth from an Insect’s Point of View.” Id.  
179 See A Triarchic Analysis of an Aptitude-Treatment Interaction, in THE ESSENTIAL STERNBERG, supra note 176, at 220 (describing the Sternberg Triarchic Abilities Test).  
180 The Rainbow Project, supra note 21, at 322.  
181 Id. at 323.  
182 Id.
The study of the “Rainbow measures”—new tests of the three intelligences conceived as constituting intelligence under the triarchic theory—was based on data collected at fifteen schools across the United States.\(^{183}\) It reported the results of regression analysis of how much of the variation in test-takers’ college GPA—college scholastic performance—was explained when the triarchic measures were used in addition to high school GPA and SAT score and when the new tests were used in lieu of the SAT.\(^{184}\) The triarchic tests alone were shown to have twice the practical predictive power of the SAT alone—“approximately double the predicted amount of variance in college GPA when compared with the SAT alone (comparative \(r^2\) values of .199 to .098 respectively)\(^{185}\)—and explained variation in college GPA unexplained by high school GPA and the SAT. Significantly, the Rainbow Project study also found that “the triarchic tests appear to reduce race and ethnicity differences relative to traditional assessments of abilities such as the SAT.”\(^{186}\) The study observed that:

> [T]he triarchic measures predict an additional 8.9% of college GPA beyond the initial 15.6% contributed by the SAT and high school GPA. These findings, combined with the substantial reduction of between-ethnicity differences, make a compelling case for furthering the study of the measurement of analytical, creative, and practical skills for predicting success in college.\(^{187}\)

The major finding of the Rainbow Project study is that the \(r^2\) value of .199 of the systems-based triarchic tests shows these new tests explaining 19.9% of the variation in the college GPA of the test-takers. This is a significant increase over the 9.8% of the variance in college GPA accounted for by the SAT and high school GPA.\(^{188}\)

---

\(^{183}\) See id. at 326–28. The Rainbow measures of analytical, practical, and creative abilities used in the Rainbow Project were the Sternberg Triarchic Abilities Test (“STAT”)—a multiple choice test developed as a means of capturing test-takers’ analytical, practical, and creative skills—and open-ended performance tasks designed to measure test-takers’ creative and practical skills. In addition to the creative skills measured by the STAT, creative tasks require test-takers to respond to a choice of topics or stimuli on which to base a creative oral or written story or cartoon caption. Practical tasks are designed to measure test-takers’ “ability to acquire useful knowledge from experience, including ‘tacit knowledge’ that is not explicitly taught and is often difficult to articulate, and to apply [tacit] knowledge to solving complex everyday problems.” Id. at 328. Again, in addition to the STAT measurement of practical skills, the Rainbow Project tests included performance measures of practical skills—three “situational judgment inventories”[: the Everyday Situational Judgment Inventory (Movies), the Common Sense Questionnaire, and the College Life Questionnaire, each of which was designed and shown to measure different types of “tacit knowledge” understood to be a central characteristic of intelligent behavior under the triarchic theory of successful intelligence. Id. at 329.

\(^{184}\) Id. at 329, 344.

\(^{185}\) Id. at 343–44.

\(^{186}\) Id. at 342–43.

\(^{187}\) Id. at 344.
grades \[ r^2 \text{ value of .098} \] explained by the conventional factorist SAT test. 188 Noting that racial group differences in test scores are not reduced completely by the triachic tests, it is Sternberg’s conclusion that the Rainbow Project regression studies suggest valid and reliable tests “can be designed that reduce ethnic and socially defined racial group differences on standardized tests, particularly for historically disadvantaged groups such as blacks and Latinos.” 189 Sternberg also notes that “[t]hese findings have important implications for reducing adverse impact in college admissions.” 190

IV. IMPLICATIONS OF MORE INTELLIGENT TESTS

[T]here is much to be said for the view that the use of tests and other measures to ‘predict’ academic performance is a poor substitute for a system that gives every applicant a chance to prove he can succeed in the study of law. The rallying cry that in the absence of racial discrimination in admissions there would be true meritocracy ignores the fact that the entire process is poisoned by numerous exceptions to ‘merit’ . . . . [E]xceptions to ‘true’ meritocracy give the lie to protestations that merit admissions are in fact the order of the day at the Nation’s universities.

—Justice Thomas in Grutter v. Bollinger 191

Intelligence research findings that tests designed according to multi-dimensional theories of intelligence do a better job of predicting the future college performance of test-takers than traditional factorist tests call into question the presumption that the individuals with the highest factorist test scores possess greater admissions-related merit than low scorers with similar grades and other credentials. The major implication of the intelligence theory research described in this Article is that it suggests the need to rethink the role that factorist tests currently play in selective admissions. If educational institutions rely on traditional factorist tests like the SAT without also using broader and more complete tests such as Sternberg’s triarchic tests, rejected applicants can point to research like Sternberg’s as empirical evidence that selective universities and graduate institutions are not using the most predictive, least racially skewed test available; these individuals could assert that a university’s reliance on inferior mental testing technology—tests found to be less predic-

188 See id. at 342.
189 Id. at 344.
190 Id.
tive of test-takers’ future grades—is an unfair and anti-meritocratic basis for comparing applicants.

Assessing applicants based on scores on tests premised on an incomplete or flawed definition of intelligence denies them a fair opportunity to compete. In the context of tests that have a racially disparate impact on the basis of race, the test user that receives federal funding is obligated to demonstrate the “educational necessity” of the tests in question.192 Showing that the use of the tests is an educational necessity requires educational institutions receiving federal funds to demonstrate that their selection criteria (admissions tests) are accurate measures of admissions-related merit and that no less discriminatory alternative measures are available.195

A. Prior Legal Disputes Related to Race, Merit, and Testing

The most contentious aspects of the debate over the proper role of standardized tests in assessing admissions-related merit is the “racial impact” of relying on tests as an admissions criterion. Lay and legal critics have long argued that the racial and eugenic views of the men who designed the first g-based tests should be considered as evidence that their modern progeny—tests like the SAT—have inherent racial, ethnic, and class biases built into their structure.194 Yet, such arguments have not ruled the day. Instead, it is the predictive power of tests like the SAT—the attribute that psychometricians dating back to Terman and Brigham identified long ago as noteworthy—that con-

192 West-Faulcon, supra note 5, at 1123–24.
193 See id. at 1125–28 (discussing how universities must show the use of admissions criterion to be an “educational necessity”). In fact, the intelligence research that has been the focus of this Article suggests better, less racially skewed tests could make affirmative action unnecessary one day in light of tests without (or with substantially smaller) racial differences in scores. Granted, it could be said that selective universities could and should, in light of modern intelligence research, stop relying on factorist tests altogether. It could be argued that race-conscious action to ameliorate racially differential deficiencies in the measurement capacity of admissions tests does not trigger strict scrutiny or constitute a compelling justification to consider race under the Fourteenth Amendment.
194 See, e.g., Derrick Bell, Diversity’s Distractions, 103 COLUM. L. REV. 1622, 1630–31 (2003) (pointing to studies showing that admissions tests do not predict performance, but “measure quite accurately the incomes of the applicants’ parents”). It is even argued that the entire enterprise of mental testing operates to reinforce various types of social hierarchy, including, but not limited to, racial hierarchy. See JEAN-CLAUDE CROIZET, The Racism of Intelligence: How Mental Testing Practices Have Constituted an Institutionalized Form of Group Domination, in HANDBOOK OF AFRICAN AMERICAN CITIZENSHIP (Henry L. Gates ed., forthcoming Sept. 2011) (reviewing the history of mental intelligence testing and arguing that “concepts of merit and intelligence have played a major role as control ideologies in sustaining the long-term expropriative relationship between Blacks and Whites”).
continues to be the strongest argument for continued reliance on g-based tests as admissions criteria.\footnote{In the specific context of college admissions tests, the asserted predictive power is that test-takers’ scores on admissions tests correlate to the test-takers’ future academic achievement.}

That differences in racial group averages in mental test scores exist combined with the fact that there are many competing explanations for these well documented “racial gaps” in traditional mental test scores are major reasons why the use of mental testing in selective admissions has engendered substantial controversy and prompted numerous legal and policy challenges. Such complaints can take multiple forms. Rejected white applicants may accuse selective universities of “reverse discrimination” for admitting lower-scoring non-whites in violation of the Fourteenth Amendment Equal Protection Clause and federal civil rights statutes\footnote{See, e.g., Grutter, 539 U.S. at 343–44 (upholding the University of Michigan Law School’s admissions decision-making process as not in violation of the Equal Protection Clause of the Fourteenth Amendment); Gratz v. Bollinger, 539 U.S. 244, 251 (2003) (holding the University of Michigan’s undergraduate admissions test to be in violation of the Equal Protection Clause of the Fourteenth Amendment); Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265, 320 (1978) (holding the petitioner’s special admission program, which admitted lower-scoring minority applicants, was unconstitutional and in violation of the Equal Protection Clause of the Fourteenth Amendment); DeFunis v. Odegaard, 416 U.S. 312, 319–20 (1974) (allowing student to make a reverse discrimination claim under the Equal Protection Clause of the Fourteenth Amendment but declining to hear the merits of the case because petitioner would be allowed to complete his final term and graduate from the law school); see also Smith v. Univ. of Wash., 392 F.3d 367, 376 (9th Cir. 2004) (holding that the law school’s admissions program did not violate the Fourteenth Amendment); Hopwood v. Texas, 78 F.3d 932, 934 (5th Cir. 1996) (striking down the Texas School of Law’s affirmative action admissions program because it was in violation of the Fourteenth Amendment).} or, in states with anti-affirmative action laws,\footnote{See West-Faulcon, supra note 5, at 1086 (discussing California, Washington, Michigan, and Nebraska state anti-affirmative action laws).} of operating “under-the-table affirmative action” policies that violate state laws prohibiting racial “preferences” by public universities.\footnote{Id. at 1104 n.90.} Rejected non-whites from racial groups at the low end of the racial gap in conventional test scores, such as African Americans, Latinos, and Filipinos,\footnote{See cases cited supra note 196.} have contended that reliance on conventional mental tests as selective admissions criteria results in unjustified disparate impact that violates federal civil rights law such as Title VI of the Civil Rights Act of 1964.\footnote{See, e.g., Complaint at 3, Rios v. Regents of the Univ. of Cal., No. 99-0525 (N.D. Cal. 1999), settled sub nom. Castaneda v. Regents of the Univ. of Cal., No. 99-0525 (N.D. Cal. 2003) (citing a claim brought by Latino, African American, and Filipino American college app-
cial groups at the high end of the racial gap in conventional test scores, such as Chinese, Japanese, Korean, and some South Asian racial groups, object to changes in selective admissions policies that reduce the amount of weight given to applicants’ scores on mental tests like the SAT as racially motivated attempts to decrease the number of Asian Americans at particular elite universities. 201

Racial differences in admissions test scores have triggered numerous lawsuits over the past four decades. For the most part, the plaintiffs in such lawsuits have been rejected white applicants challenging the affirmative action admissions policies of selective universities. In those cases, white plaintiffs relied on numerical differences between their tests scores and the test scores of admitted non-white applicants—specifically, that their higher test scores should be understood as scientific proof of their greater admissions-related merit than lower-scoring non-white students—to argue that race-based affirmative action violated their federal civil rights. 202

On the other hand, critics of testing allege that reliance on SAT scores is undemocratic 203 and unfairly reinforces racial and socioeconomic hierarchies 204 that disproportionately harm African-American and Latino applicants due to those groups’ lower averaged SAT scores. In addition, critics point to research showing that SAT scores

---


202 The Supreme Court’s Fourteenth Amendment equal protection jurisprudence as well as Fourteenth Amendment equal protection legal scholarship over the past four decades has also been focused on such claims. See, e.g., RONALD DWORKIN, A MATTER OF PRINCIPLE 293-315 (1985) (discussing Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265 (1978)); Lani Guinier, Comment, Admissions Rituals as Political Acts: Guardians at the Gates of Our Democratic Ideals, 117 HARV. L. REV. 113, 117 (2003) (discussing Grutter v. Bollinger, 539 U.S. 306 (2003)); Cheryl I. Harris, What the Supreme Court Did Not Hear in Grutter and Gratz, 51 DRAKE L. REV. 697, 703–707 (2003) (same); Lawrence, supra note 44, at 935–34 (same). All of these authors discuss cases in which rejected white applicants claim universities’ consideration of race as an admissions criterion violates the Fourteenth Amendment Equal Protection Clause and Title VI of the Civil Rights Act of 1964.

203 See Guinier, Admissions Rituals, supra note 202, at 116 (“At the same time that higher education is considered a democratic and educational necessity to many, it remains beyond the reach of all but a Few.”).

204 See Lawrence, supra note 44, at 972 (arguing for a re-examination of “merit in light of the university’s commitment to the goal of fighting racism”); Daria Roithmayr, Deconstructing the Distinction Between Bias and Merit, 85 CAL. L. REV. 1449, 1452 (1997) (discussing the “radical critique of merit,” which “argues that merit standards disproportionately exclude people of color and women because the standards historically have been developed by members of dominant groups in ways that end up favoring them”).
are less useful in predicting the future academic performance of applicants from certain racial groups. In fact, proponents of race-based affirmative action often point to the academic and overall postgraduate success of racial affirmative action as proof that some of the most successful minority alumni would have been denied admission if selective universities admitted strictly based on rank-order test scores.

Nevertheless, arguments that g-based factorist tests are biased against particular groups have been trumped by statistical evidence that conventional factorist tests do, in fact, predict test-takers’ first-year grades to some extent. The predictive capacity of g-based tests has, from the earliest days of mass mental testing, been the bulwark of psychometric theory. Accordingly, the longstanding policy and legal justification for relying on traditional factorist tests like the SAT has been that such tests, despite their disproportionate exclusion of particular non-white racial groups like African Americans, Latinos, and Native Americans, are fair proxies for admissions-related merit because they have been scientifically demonstrated to predict test-takers’ future academic success.

Previously, the major empirical critique of conventional factorist tests has been that the substantial racial and socioeconomic differences in factorist test scores do not predict the future grades of test-takers (white and non-white) with sufficient precision to counterbal-

205 See Gregg Thomson, Is the SAT a “Good Predictor” of Graduation Rates? The Failure of “Common Sense” and Conventional Expertise and a New Approach to the Question 8 (Dec. 1998) (unpublished manuscript) (on file with author) (presenting data refuting popular notions that SAT scores strongly predict graduation rates and that African Americans admitted with lower scores have lower graduation rates).

206 See, e.g., Luke Charles Harris & Uma Narayan, Affirmative Action and the Myth of Preferential Treatment: A Transformative Critique of the Terms of the Affirmative Action Debate, 11 HARV. BLACKLETTER L.J. 1, 22 (1994) (“The literature on such standardized tests demonstrates that they are inaccurate indicators even with respect to their limited stated objective of predicting students’ first-year grades in college and professional school.”); Richard O. Lempert et. al., Michigan’s Minority Graduates in Practice: The River Runs Through Law School, 25 L. & SOC. INQUIRY 395, 496 (2000) (“LSAT scores and UGPAs, the admissions credentials that the opponents of law school affirmative action would privilege for their supposed bearing on ‘merit’ and ‘fitness to practice law,’ bear for one school’s graduates little if any relationship to measures of later practice success and societal contribution.”); see also Lani Guinier, Comment, Confirmative Action, 25 L. & SOC. INQUIRY 565, 565 (2000) (relying on a study of the careers of non-white Michigan Law School graduates to argue that “conventional test-based admission policies both mask and support deep flaws in the way we allocate opportunity and privilege” by selecting applicants “who then often fail to give back to society” and failing “to identify those who in fact have much more to give and do give in service of the [legal] profession and its larger goals”).

207 See supra Part II.B.

208 See id.
ance the clear statistical evidence that factorist tests are highly correlated to the test-takers’ parents’ income and education. Examples of critics of the role that conventional tests currently play in selective admissions include Lani Guinier and even Justice Clarence Thomas. In general terms, such critiques question the predictive capacity of traditional mental tests and disagree with testing experts as to the significance of the incremental predictive power of factorist tests. In addition to pointing out the predictive limitations of conventional admissions tests, legal scholars like Charles Lawrence have challenged the objectivity of merit-based standards in selective admissions and suggested alternative bases for allocating access to selective higher education that dismantle racial and socioeconomic hierarchy.

This Article is an intervention of a different sort. Empirical evidence that systems-based tests are more predictive than conventional factorist tests with less racially disparate impact does more than demonstrate the need to improve upon the predictive power of conventional tests. If conventional admissions tests are demonstrably less predictive, or “less intelligent” tests, it becomes possible for rejected applicants to argue such tests are no longer the best proxies for admissions-related merit because they make a demonstrable racial difference in admissions outcomes that could be avoided using admissions tests based on broader theories of intelligence.

This fundamentally changes the legal and policy analysis of whether a selective university admissions policy that relies on factorist tests is fair to non-white test-takers like African Americans and Latinos. It presents a new “test deficiency” explanation for racial differences in group averaged scores on traditional factorist tests—that the “g” general factor intelligence theory upon which factorist tests are premised is inferior to new contemporary theories of intelligence. Hence, this argument continues, beyond there being substantial

209 See Sturm & Guinier, supra note 128, at 974 (recognizing “that high school grades are more predictive of college freshman-year grades than the SAT”); see also supra Part I.

210 See Lawrence, supra note 44, at 930 (writing “as an unambivalent advocate for affirmative action”); Harris & Narayan, supra note 206, at 25 (“[I]n circumstances in which a person’s class, race, or gender operates as a group status that impedes equality of opportunity, we suggest that affirmative action policies promote a greater degree of equality of opportunity than would otherwise be afforded to members of these groups.”); see also Roithmayr, supra note 204, at 1453 (“[M]erit standards are necessarily the effect of subjective, social and contingent race-conscious preferences for particular kinds of abilities.”). Guinier’s scholarship might properly be placed in this category as well. See, e.g., Guinier, Confirmative Action, supra note 206.

211 In contrast to g-based theories, systems-based theories like the triarchic theory reject the hereditarian theory that intelligence is distributed disproportionately and genetically to different racial groups.
“room for improvement” in the predictive power of traditional factorist admissions tests like the SAT, there is scientific support for the conclusion that true racial differences in test-measured mental ability are smaller than the more racially skewed scores of g-based factorist tests.

B. Resolving the Tension Between Merit and Diversity

The empirical evidence that new non-factorist tests do a better job of predicting test-takers' future academic performance with smaller racial group differences in test scores demonstrates the error of presuming racial diversity and academic merit are impossible to achieve simultaneously. The assumption that applicants' mental ability and likelihood of future academic success align in the same rank order as their test scores is based partly on intuition and partly on statistical and empirical evidence. The perception that, all else being equal, the rank-order relationship between test-takers' scores is a fair measure of admissions-related merit is the basis upon which universities are often criticized for rejecting applicants with high scores in favor of lower-scoring applicants. Often, the consideration of non-test, non-grade criteria is derided as a deviation from the fair application of standards of academic merit, whereas consideration of standardized test scores, along with prior grades, is deemed a proper application of conventional admissions-related merit standards.\footnote{See, e.g., Daniel A. Farber & Suzanna Sherry, Beyond All Reason: The Radical Assault on Truth in American Law 13 (1997) (defending “the accepted ideal of making decisions on the basis of merit,” but “not all the specifics of how decisions are currently made”).}

As a consequence, the scientific rationale for treating a higher test score as a proxy for greater admissions-related merit generally receives minimal scrutiny.\footnote{By contrast, selective educational institutions’ use of “soft,” non-test, non-grade variables and applicants’ race as factors in admissions are perceived and challenged as driven only by ad hoc institutional preferences for particular non-academic traits, instead of a scientifically defensible theory. The problem is exacerbated because universities typically lack, or at least fail to offer, an empirical basis for deviating from exclusive reliance on test scores and grades as the sole basis for their admissions decisions. University leaders and admissions officials typically fail to offer a convincing rationale for the consideration of “soft” variables at all and, if they do, they are reluctant to share details as to the manner and degree to which their institutions rely on “soft” versus “hard” variables. Lani Guinier and others have critiqued this lack of “transparency” in selective higher education admissions. See, e.g., Guinier, Admissions Rituals, supra note 202, at 188.} Without an explicit articulation of a rationale for deviating from them, applicant scores on mental tests like
the SAT and, to a lesser extent, grades, are presumed to be scientific proof that the individual with the highest numbers has the greatest admissions-related merit. Non-test, non-grade admissions variables such as the extent to which an applicant has “overcome adversity” are regularly criticized as deviations from merit principles that encourage “narratives of suffering” instead of merit-based selection. Likewise, based on similar grounds, race-based affirmative action in selective admissions is intensely criticized for resulting in both the racially inconsistent application of test score standards and the rejection of white applicants with higher scores than some admitted minority applicants.

As using standardized tests to allocate resources in our society becomes increasingly common, the need is heightened for an analytical framework for evaluating the fairness and legality of particular uses of

214 Because the range and ability to draw distinctions on the basis of test scores is greater and because of the implicit presumption that test scores are scientifically valid measures of intellectual ability and the racially disparate impact associated with reliance on tests rather than on reliance on grades, charges that someone with a lower test score was admitted at the expense of someone with a higher test score is perceived as more unfair than a similar assertion based on a difference in grades.


216 Rejected white applicants have alleged frequently that their constitutional and statutory civil rights were violated because non-whites with lower test scores were admitted. See, e.g., Gratz v. Bollinger, 539 U.S. 244, 249 (2003); Grutter v. Bollinger, 539 U.S. 306, 311 (2003); Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265, 269–70 (1978); DeFunis v. Odegaard, 416 U.S. 312, 314 (1974); Smith v. Univ. of Wash., 392 F.3d 367 (9th Cir. 2004), cert. denied, 546 U.S. 813 (2005); Hopwood v. Texas, 78 F.3d 932 (5th Cir. 1996). The now classic objection to race-based affirmative action is that race is an improper basis for distinguishing amongst individuals and that it unjustly burdens or disadvantages individual applicants in a selection process when those particular individuals are not responsible for the racial wrongs that affirmative action is designed to remedy. Likewise, opponents of race-based affirmative action criticize universities for considering race on the grounds that racial group affiliation is irrelevant to individual merit and, thereby, an improper consideration in affording access to elite higher education. See Antonin Scalia, Commentary, The Disease as Cure: “In Order to Get Beyond Racism, We Must First Take Account of Race”, 57 WASH. U. L. Q. 147, 147 (1979) (stating, in reference to Bakke, “I have grave doubts about the wisdom of where we are going in affirmative action, and in equal protection generally”). Opponents also contend that affirmative action inflicts “stigmatic” and performance harms on its beneficiaries: it causes non-beneficiaries to resent and underestimate members of racial groups that are typically beneficiaries and suppresses beneficiaries’ future academic and professional success. See, e.g., STEPHEN L. CARTER, REFLECTIONS OF AN AFFIRMATIVE ACTION BABY (1991). But despite the many objections and policy arguments against affirmative action, the U.S. Supreme Court has held that “narrowly-tailored” race-conscious consideration of race in admissions used for the purpose of increasing an undergraduate and graduate educational institution’s multi-faceted “diversity” is constitutional. See, e.g., Grutter, 539 U.S. at 343–44; Bakke, 438 U.S. at 320.
tests. Here, I suggest that a useful approach is to conceive of mental tests as technology—neither inherently evil nor inherently infallible. Evidence that a test user has adopted an inferior form of testing technology is logically and, under Title VI disparate impact theory, legally relevant to challenging or justifying a particular use of a test.

If tests like the multi-dimensional triarchic admissions tests are more accurate predictors of future grades than conventional tests, a new question is raised about the fairness of the role of factorist tests in selective higher education admissions—whether basing admissions on factorist test scores unfairly distorts the true admissions-related merit of individual applicants and racial groups. If factorist tests are inferior in predictive effectiveness to systems tests, there are potential legal and policy implications that flow from a university’s failure to either abandon factorist tests or to augment them with more predictive tests.

The core of this argument—that the legal concept of lodging a “test deficiency” administrative complaint under Title VI\textsuperscript{217} or defending against allegations of reverse discrimination under the Equal Protection Clause—is not based on the grounds that traditional $g$-based factorist tests are “culturally biased.” Instead, this approach posits that there is both practical and legal significance to the fact that some alternatives to factorist tests—tests designed according to broader definitions of intelligence—are more predictive of test-takers’ future academic performance and result in smaller racial differences in test scores. Correcting for a demonstrated tendency of factorist tests to make racially skewed errors in assessing admissions-related merit operates to adhere to, not deviate from, academic merit-based admissions. In fact, race consciousness for the purpose of compensating for test deficiency could be sufficient to establish a legal justification for the explicit consideration of race in selective admissions.\textsuperscript{218}

\textsuperscript{217} See West-Faulcon, supra note 5, at 1123–24 (discussing such complaints). In Alexander v. Sandoval, 532 U.S. 275 (2001), the Supreme Court ruled that no private right of action exists to enforce Title VI disparate impact regulations. Whether rejected minority applicants may bring Title VI effect-discrimination cases under 42 U.S.C. § 1983—an approach endorsed by Justice Stevens in his dissent in Sandoval—has not been decided definitively by the Court. However, even if private enforcement of Title VI regulations is precluded, individuals are still permitted to file complaints with the U.S. Department of Education OCR alleging that an institution’s admissions policies have a Title VI discriminatory effect on the basis of race. See supra note 5.

\textsuperscript{218} Universities are uniquely situated, under the Court’s jurisprudence, and have been found to have substantial academic freedom to develop institution-specific conceptions of admissions-related merit. Rejecting $g$-centered definitions of intelligence in favor of systems theories such as Sternberg’s triarchic theory would arguably be within the First Amendment academic freedom of selective universities. Cf. Bakke, 438 U.S. at 320 (“[T]he courts
A new generation of “more intelligent” standardized tests provides a potential rationale for affirmative action—the test deficiency rationale—that relies on the traditional conception of merit that has previously been used only to charge that affirmative action is anti-meritocratic and unfair. The fact that more refined and more predictive intelligence and college admissions tests result in less racially disparate impact on minority test-takers opens a path to an outcome that universities have long characterized as impossible: merit-based diversity.

C. “Test Deficiency” as Complaint and Defense

Several legal and policy implications stem from universities’ federal constitutional and statutory obligations to avoid racial discrimination in selective admissions. More intelligently designed tests may generate civil rights administrative complaints alleging Title VI disparate impact discrimination. Rejected non-white applicants can argue that they would be admitted in greater numbers if tests that were both more predictive and less racially discriminatory were used. Additionally, these more intelligently designed testing instruments may provide a new, more enduring legal defense of race-based affirmative action and may also constitute an empirically supported policy justification for a university’s reliance on non-test score admissions criteria, such as the non-test score admissions variables often considered as part of holistic admissions policies.

The theoretical framework for this argument is drawn from Justice Powell’s decision in Regents of the University of California v. Bakke and below failed to recognize that the State has a substantial interest that legitimately may be served by a properly devised admissions program . . . .”

See supra Part I.

220 The predictive limitations of g-based tests and the traditional factorist theory of intelligence may provide the basis for a new doctrinal approach to affirmative action in higher education admissions. As suggested by Justice Powell in footnote forty-three to his controlling opinion in Bakke, the need to “cure” inaccuracies in a test’s predictive ability not only constitutes a compelling interest under the Court’s equal protection jurisprudence but is arguably subject to either intermediate scrutiny or rational basis review. In Bakke, Powell observed:

Racial classifications in admissions conceivably could serve a fifth purpose, one [not articulated by UC Davis Medical School] . . . : fair appraisal of each individual’s academic promise . . . . To the extent that race and ethnic background were considered only to the extent of curing established inaccuracies in predicting academic performance, it might be argued that there is no “preference” at all.

Bakke, 438 U.S. at 306 n.43 (emphasis added). Powell’s view is that using race to correct for test inaccuracies is not a preference. This is essentially an articulation of the test deficiency rationale I identify here.
echoed by Justice Kennedy in *Ricci v. DeStefano*\(^{221}\); fairness in selection prohibits reliance on tests that inadequately measure individual test-takers’ merit if better tests with smaller racial differences in scores are available. The new “test deficiency” defense I propose here could augment and bolster the “diversity rationale” invoked by selective universities sued by rejected white applicants alleging their rejection constituted illegal race discrimination. Presuming that multidimensional theories of intelligence and college admissions tests will not eliminate the need for affirmative action altogether, universities might employ such theories and the decrease in racial disparities associated with such tests to re-shape the affirmative action debate. This Article articulates the theoretical foundation for the introduction of test deficiency into the doctrinal framework of the Court’s

---

\(^{221}\) Justice Kennedy’s conclusion in *Ricci v. DeStefano* that it is permissible to consider race to increase a test’s overall fairness—such as pre-test administration race-conscious undertaken to design tests that are more fair to test-takers of all races—also supports this conclusion. See *Ricci v. DeStefano*, 129 S. Ct. 2658, 2677 (2009) (holding that, to engage in intentional discrimination, an employer “must have a strong basis in evidence to believe it will be subject to disparate-impact liability if it fails to take the race-conscious, discriminatory action”). But see *Parents Involved in Cmty. Schs.*, 551 U.S. 701, 782–84 (2007) (Kennedy, J., concurring) (arguing that any government use of individual racial classifications is subject to strict scrutiny). Even if the Court, as is likely given its current composition, were to take the more probable approach of holding that a university’s consideration of race to correct for test deficiency does trigger strict scrutiny, a new “test deficiency” rationale for affirmative action in higher education is a plausible alternative doctrinal approach to evaluating the constitutionality of race-based affirmative action in higher education admissions. As has been often true for the current Court, Justice Kennedy’s position would likely be central to the outcome of a case in which the plaintiffs or defendants relied upon test deficiency.

Although the *Ricci* case addressed the evidentiary burden employers must satisfy before taking post-test-administration actions to correct for a test’s racially disparate impact, Kennedy’s majority opinion in *Ricci* suggests that race-consciousness, if done prior to administration of the test warrants different treatment. See Cheryl I. Harris & Kimberly West-Faulcon, *Reading Ricci: Whitening Discrimination, Racing Test Fairness*, 58 UCLA L. REV. 73, 102 (2010) (explaining the Court’s holding in *Ricci* and arguing that *Ricci* misinterprets consciousness of race for making hiring decisions directly based on race). In the *Ricci* decision, Kennedy suggests that race-conscious action taken to design “more intelligent” tests should be subject to a different, less stringent, standard than race-consciousness after a “less intelligent” test has been administered. He writes:

> Title VII does not prohibit an employer from considering before administering a test or practice, how to design that test or practice in order to provide a fair opportunity for all individuals, regardless of their race. And when, during the test-design stage, an employer invites comments to ensure the test is fair, that process can provide a common ground for open discussions toward that end. *Ricci v. DeStefano*, 129 S. Ct. 2658, 2677 (2009). This language may mean Justice Kennedy would potentially endorse race-consciousness for the purpose of designing a test that “provide[s] a fair opportunity for all individuals, regardless of their race,” a view similar to Justice Powell’s observation that “curing established inaccuracies in predicting academic performance” is a legitimate justification for considering race as a factor in selection admissions. *Bakke*, 438 U.S. at 306 n.43.
Fourteenth Amendment constitutional jurisprudence, Title VI statutory disparate treatment analysis, and disparate impact analysis under Title VI implementing regulations.

Universities that rely on race to correct for the predictive limitations of traditional factorist tests like the SAT could invoke the use of race to correct for deficiencies in the predictive capacities of the tests—as a necessary requisite to ensuring fairness in individual competition for selective admission. Social science research demonstrating that new systems-based college admissions tests are twice as predictive as factorist tests like the SAT is the type of empirical evidence of “test deficiency” that could potentially justify the consideration of race in selective admissions.  

1. Title VI Test Deficiency

Rejected minority applicants from racial groups at the low end of the racial gap in conventional test scores, such as African Americans and Latinos, may contend that reliance on conventional mental tests as selective admissions criteria results in unjustified disparate impact that violates federal civil rights law, such as Title VI of the Civil Rights Act of 1964. In particular, rejected non-white test-takers who are members of racial groups at the low end of the conventional racial test score gap might reasonably take the legal position that the racial gap in factorist test scores is, at least in part, attributable to the predictive deficiencies of factorist tests as compared to multi-dimensional tests. Accordingly, such rejected African-American and Latino applicants could file Title VI disparate impact complaints with the Department of Education Office of Civil Rights alleging that racial disparities in admissions rates stemming from the use of admissions tests premised on a flawed theory of intelligence cannot be justified as an “educational necessity.”

If new multi-dimensional admissions tests have been empirically proven to be more predictive with smaller racial differences in test scores, rejected applicants making Title VI disparate impact claims may have a decent chance of proving the existence of “less discriminatory alternatives” to conventional factorist tests. Thus, under Title VI disparate impact theory, proof of test deficiency could expose test users—selective universities and graduate institutions—to potential

222 See id. (stating that racial classifications could serve the purpose of “curing established inaccuracies in predicting academic performance”).

223 See generally West-Faulcon, supra note 5, at 1099 (noting that private Title VI disparate impact complaints can no longer be litigated in federal court after the Sandoval decision).
Title VI disparate impact liability. To immunize and defend themselves against Title VI test deficiency charges by rejected minority applicants and against challenges by rejected white applicants, universities could either begin to use systems-based admissions tests or to consider race and other non-test scores as factors for the purpose of correcting the prediction errors that result from reliance on factorist tests.

2. Reverse Discrimination Equal Protection Challenges

Another implication of this analysis is that universities may have a promising test deficiency policy justification for adopting admissions criteria that potentially compensate for the flaws in the intelligence theory that undergirds factorist tests. Universities have a possible test deficiency legal justification for correcting for the fact that factorist tests have been found to have greater racial group differences in scores than systems-based admissions tests yet do a worse job of predicting the future academic performance of test-takers. Test deficiency is a potential justification for choosing not to rely on tests such as the SAT or mitigating the extent of that reliance in particular circumstances. In addition, evidence that tests like the SAT have less predictive power than systems tests, such as triarchic admissions tests designed according to the “theory of successful intelligence,” also makes it less likely that rejected white applicants can demonstrate that inconsistent reliance on factorist test scores constitutes “reverse discrimination” in violation of the Fourteenth Amendment and federal statutes prohibiting race discrimination such as Title VI.

Empirical evidence of the scientific deficiencies of factorist tests weakens race discrimination claims by rejected white applicants who would otherwise rely on the predictive power of factorist tests as evidence that they possess greater admissions-related merit than non-whites admitted with lower test scores. A central, though not necessarily explicit, premise of rejected white applicants’ “reverse discrimination” claims is usually that tests like the SAT, GRE, LSAT, and MCAT are based on sound scientific theory and, as a result, accurate-

---

224 Title VI disparate impact law requires federally funded educational institutions to justify the use of admissions tests that have a racially disproportionate impact as constituting an “educational necessity.” See supra notes 192–93 and accompanying text. This requirement is analogous to the “business necessity” requirement imposed on employers by Title VII of the Civil Rights Act of 1964. Even institutions able to demonstrate an educational necessity for relying on tests that have a racially adverse impact violate Title VI disparate impact law if it can be shown that less discriminatory alternative tests are available.

225 See supra Part III.C.
ly sort and rank individuals according to their mental ability. The smaller racial differences in scores on more intelligently designed admissions tests undermines the scientific credibility of claims that having a relatively higher factorist test score makes an applicant per se more deserving—more entitled to selection—than an otherwise similarly situated but lower-scoring applicant.

As such, the existence of “more intelligent” tests with smaller racial differences in scores shifts the legal terrain in conflicts involving race and testing. The predictive inadequacies and scientifically unjustified racial differences in scores of conventional factorist tests like the SAT may be legally cognizable “test deficiencies.” Specifically, I suggest that Fourteenth Amendment equal protection jurisprudence and Title VI disparate impact standards may be structured to recognize the concept of test deficiency that is the focus of this Article. It is on this basis that I assert that modern innovations in intelligence theory and test development offer an empirical basis for rejected minority applicants to raise the test deficiencies of factorist tests as a legal complaint and for educational institutions to raise the same test deficiencies as a legal defense in cases where rejected whites allege reverse discrimination. Hence, modern intelligence research potentially strengthens race discrimination claims by rejected minority applicants who allege factorist tests are “deficient” and weakens race discrimination claims by rejected white applicants relying on the predictive power of factorist tests to challenge their rejection by selective universities.\footnote{See id. at 1109–11 (describing test deficiency rationale).}

D. Impediments to More Intelligent Testing

Nevertheless, it is likely that selective universities will identify numerous potential practical problems and impediments to their adoption and use of more intelligently designed tests in real-world admissions. First and foremost, they may argue that because these tests have yet to be produced on a mass scale, it is improper to consider them a less discriminatory alternative under Title VI disparate impact analysis. In fact, some institutions will also likely point out that there remains ongoing debate within the intelligence research community as to how to properly define intelligence and are likely to rely on the expert opinions of adherents to the g-oriented psychometric view to criticize newly designed tests as measuring something other than true intelligence. Institutions might also argue that newly designed tests
unfairly exclude and discriminate against certain groups, particularly higher-scoring racial groups like whites and Asian Americans, who, on average, score higher on conventional factorist admissions tests. Universities that continue to rely on factorist test scores may also criticize new tests like Sternberg’s triarchic measures for being just as coachable as, or potentially even more coachable than, conventional g-oriented factorist tests such that racial differences in scores emerge on new tests that mirror the racial gaps in conventional factorist test scores. Another response by selective universities may be to argue that other scientific research already calls, or may at some point in the future call, into question the scientific research that is the focus of this Article—research may challenge whether new tests, like Sternberg’s, are actually better and less racially skewed.

While it is true that any of the above mentioned practical concerns could impact selective universities’ decisions to eliminate their use of factorist tests or their decisions to adopt new, more intelligently designed tests, the theoretical and legal implications of modern intelligence research are still extremely significant. There are still implications for considering the role of tests in America’s meritocracy ideology and, even if scientific questions remain unanswered or even if it is decades before more intelligent tests are operational, I contend there are current legal implications raised by the scientific insights arising from the ongoing debate within the field of intelligence research over the accuracy of theories of intelligence and the theoretical underpinnings of factorist g-oriented tests. While other implications warrant exploration, the greater risk of Title VI discriminatory effect administrative complaints filed by non-white applicants and decreased vulnerability to Fourteenth Amendment and Title VI claims of purposeful discrimination by rejected white applicants are the central implications of more intelligent tests that I have identified in this Article.

CONCLUSION

The misuse of mental tests is not inherent in the idea of testing itself.

—Stephen Jay Gould

227 See, e.g., Kang, supra note 201, at 2–3 (noting that affirmative action usually results to disadvantage Asian Americans who typically score highly on the SAT and other college entrance exams).

228 GOULD, supra note 93, at 155.
As our society’s reliance on tests as merit-measuring technology becomes greater and greater, it is more critical that the law and legal scholarship consider more than the question that has dominated discourse in the past—when it may be fair to admit an individual with lower test scores over a person with higher test scores. Recognition of the limitations of the predictive power of tests and theories of intelligence dictating the design of mental tests is rarely considered outside the field of psychology and, as a result, has been insufficiently explored within legal discourse. I contend that fairness in selective admissions requires more than evaluating the fairness of selecting low scorers over high scorers; rather, it necessitates assessment of the fairness of the often implicit, sometimes explicit, conclusion that a higher scorer on a particular mental test has been scientifically proven to be more qualified than a lower scorer on the same test.

The insight of this Article’s analysis provides selective universities a much-needed theoretical basis for relying on admissions test scores or for deviating from rank-order selection based on mental test scores, as well as scientific support for admissions decisions based on better theories of intelligence. In addition, it offers a response to various accusations of bias in selection—minorities and the poor contend selective admissions are biased in favor of the racially and socio-economically privileged while rejected applicants of higher-scoring racial groups and the socio-economically privileged contend selective admissions is biased against them because it unfairly favors the downtrodden. If institutions fail to adopt new, better, less racially skewed tests, their continued reliance on g-oriented tests arguably violates Title VI implementing regulations and encourages Fourteenth Amendment reverse discrimination equal protection claims of rejected whites.

Thus, this project is a first step toward a new interdisciplinary approach to evaluating merit in selective admissions—an endeavor that has the potential to result in selection that is more fair, at the indi-
individual level, and more inclusive to all racial groups. The value of this interdisciplinary exercise is that it makes it possible to examine the legal implications of treating mental tests as technology with both limits in capacity and potential for improvement. It is also a key analytic step in beginning to bring scientifically and empirically supported theories of intelligence and measurement test theory to bear against the previously unquestioned presumption that admissions-related merit aligns in perfect rank-order with higher or lower scores on conventional factorist tests.

If the predictive limitations of conventional mental tests are due to inadequacies of \( g \)-based factorist theories of intelligence, merit-based selection in selective higher education admissions would be more fair to individuals if they were assessed based on their scores on “more intelligent” tests—tests designed according to broader and more accurate theories of intelligence. Stated another way, selection processes must take into account “test deficiencies”\(^{231}\) in order to be fair measures of merit. Under this logic, selection based on strict linear ranking of individuals’ scores on conventional \( g \)-based factorist tests without compensating for the measurement deficiencies of such tests would, contrary to what is commonly presumed, undermine conventional merit-based standards.

Evidence that traditional \( g \)-based factorist tests\(^{232}\) have been shown to have less predictive power than newly designed multi-dimensional tests destabilizes the prior presumption that relying on \( g \)-based tests in selective admissions is the linchpin holding objective standards of academic merit in place. The interdisciplinary shift I have proposed in this Article also encourages legal discourse that acknowledges the need for further innovation in mental testing—more social science research with the goal of developing tests with increased predictive power and fewer racial disparities. And, while updated testing technology is being operationalized, this Article offers a doctrinal and policy framework for evaluating the fairness and legality of relying on conventional factorist mental tests. Hence, although the central examination of this Article—the implications of scientific improvement in the intellectual merit-measuring capacity of tests—is not about race or race discrimination, there are significant racial implications that flow from this Article’s insights.

It bolsters administrative civil rights complaints lodged by minority applicants contending that reliance on conventional factorist tests

\(^{231}\) See supra Part II.
\(^{232}\) See supra Part III.C.
is not justified by “educational necessity” as required under federal antidiscrimination law. Rejected non-white applicants, particularly African Americans and Latinos, may point to the test deficiencies of traditional factorist tests as a basis for challenging selective admission policies under Title VI disparate impact law and offer new systems-based tests as a more predictive, as well as “less discriminatory alternative,” to traditional factorist tests. Additionally, evidence that tests like the SAT have less predictive power than tests designed according to the multi-dimensional “triarchic theory of intelligence” also makes it less likely that rejected white applicants will allege that inconsistent reliance on factorist test scores constitutes “reverse discrimination” that violates the Fourteenth Amendment and Title VI federal statutes prohibiting intentional race discrimination. Thus, the combination of better definitions of intelligence and more predictive tests with less racially disparate impact on minority test-takers improves the capacity of selective universities to measure applicants’ admissions-related merit and shifts the theoretical and legal terrain in conflicts involving race and testing.

Even if selective universities stop short of adopting new standardized tests modeled to measure intelligence more holistically, research substantiating systems theories of intelligence could serve as a legal as well as theoretical justification for using race and other non-test variables as corrections for the predictive inadequacies of conventional factorist tests like the SAT. If systems-based tests do a better job of identifying the most “successfully intelligent” applicants but diminish the racial disparity in test scores that has marked factorist tests historically, rejected non-white applicants and universities may convince courts that “test deficiency” explains the racial gap in SAT scores. This means that the long and bitter debate over race-based affirmative action in higher education may give way to a new generation of dialogue about race, merit, and testing in which racial diversity and merit are not deemed inherently incompatible.

Consequently, a major theoretical implication from this analysis is that it preserves the concept of meritocracy as attainable by continued innovation in testing—the possibility of more intelligently designed mental tests. Lastly and significantly, this analysis also suggests an alternative paradigm in which a more racially fair (as well as class-focused) conception of meritocracy may be articulated and eventually

---

233 See West-Faulcon, supra note 5, at 1126.
234 See The Rainbow Project, supra note 21, at 323 (“Successful intelligence is defined in terms of the ability to achieve success in life in terms of one’s personal standards, within one’s sociocultural context.”).
realized. It opens a path to an admissions outcome selective universities have traditionally argued could not be achieved: merit-based diversity.