

**THE SECOND WAY: A COMMENT ON HAWKINS AND A
CAUTIONARY NOTE**

*Richard Sander**

After a decade of intense debate, the question of whether mismatch exists has been largely resolved. Dozens of studies published in peer review journals¹ have found increasingly precise evidence of first-order mismatch effects,² and critics, who ten years ago seemed so formidable, have largely fallen silent.³

* Professor of Law, UCLA; Ph.D., Economics.

¹ A sampling includes: STEPHEN COLE & ELINOR BARBER, INCREASING FACULTY DIVERSITY: THE OCCUPATIONAL CHOICES OF HIGH-ACHIEVING MINORITY STUDENTS 100–38 (2003); Peter Arcidiacono, Esteban Aucejo & V. Joseph Hotz, *University Differences in the Graduation of Minorities in STEM Fields: Evidence from California*, AM. ECON. REV. (forthcoming 2015); Esther Duflo et al., *Peer Effects, Teaching Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya*, 101 AM. ECON. REV. 1739 (2011); Fredrick L. Smyth & John J. McArdle, *Ethnic and Gender Differences in Science Graduation at Selective Colleges With Implications for Admission Policy and College Choice*, 45 RES. HIGHER EDUC. 353, 373 (2004); Doug Williams, *Do Racial Preferences Affect Minority Learning in Law Schools?*, 10 J. EMPIRICAL LEGAL STUD. 171 (2013).

² “First-order” mismatch effects are those where there is a direct causal connection between large credential disparities among students and an observable effect. In Richard Sander, *The Stylized Critique of Mismatch*, 92 TEX. L. REV. 1637, 1642–43 (2014), I identified and discussed three types of first-order effects: learning mismatch, competition mismatch, and social mismatch. Where good measures exist, these are the easiest mismatch effects to document, precisely because of the straightforward causation path. “Second-order” effects—such as the effect of mismatch on college graduation rates—are much harder to document; thus, for example, students do not fail to graduate because of mismatch itself, but because the first-order types of mismatch may affect their success in completing a degree. Since a college may take countermeasures that affect the second-order result (i.e., by making it easier to graduate or tolerating grade inflation) without addressing the underlying first-order problem, second-order effects can be even harder to measure. See also Arcidiacono et al., *A Conversation on the Nature, Effects, and Future of Affirmative Action in Higher Education*, 17 U. PA. J. CONST. L. 695, 709–19 (2015). (Questions 4, 5, and 6)

³ So far as I know, none of the very vocal critics of law school mismatch who were writing a decade ago, such as Richard Lempert, Ian Ayres, Richard Brooks, Daniel Ho, Michelle Dauber, or David Wilkins, have followed up with successfully peer-reviewed research refuting the mismatch hypothesis. A recent joint venture by some of them known as the “Empirical Scholars Brief” has been utterly discredited. See Brief of Empirical Scholars as Amici Curiae in Support of Respondents, *Fisher v. Univ. of Tex. at Austin*, 133 S. Ct. 2411 (2012) (No. 11-345); Richard Sander, *Diversity in Legal Education and the Legal Profession: A Symposium Honoring Indiana Chief Justice Randall Shepard: Mismatch and the Empirical Scholars Brief*, 48 VAL. U. L. REV. 555, 573, 583 (2014) (discrediting the Empirical Scholars Brief).

Professor Stacy Hawkins brings to these issues a fresh perspective and a thoughtful argument.⁴ She neither agrees with nor explicitly denies the mismatch phenomenon, but contends that even if mismatch is a real problem, this merely heightens the need to both re-think the ways that colleges achieve racial diversity and to implement far-reaching changes in higher education policies. Hawkins is therefore not an apologist for the current system or someone claiming that racial preferences by themselves are the key to successful diversity. She is instead a reformer, and a fairly far-reaching reformer at that. In this sense Hawkins and I are kindred spirits; I greatly admire her spirit of reform and, when we have conversations about higher education policy in this area, we agree on a great deal.⁵

I write this piece not so much to criticize or rebut Hawkins' argument as to raise a series of strong caveats about it. Her argument is, in a way, orthogonal to the traditional mismatch debate; she finds both sides in that debate relying too much on sets of numbers—in particular, standardized test scores and grades—that she views as largely illegitimate measures of performance and merit. The task (as she sees it) is for colleges and universities to find better measures of merit and talent, and to develop strategies to nourish that talent within their academic programs. I agree with these goals, and I have spent a good deal of time working on and even implementing some of them.⁶ However, the devil is in the details. In the process of laying out her vision, Hawkins makes many assumptions about mismatch that are unwarranted and representations that are misguided. She embraces a good deal of research that, while impressive at face value, does not have as much real-world significance as she suggests, and in some cases is simply wrong. Many readers could come away from the Hawkins article concluding that things really are not so bad after all, or that colleges and universities have already gotten the message and are making the right kinds of reforms. Just the opposite is the case. Reform is vitally needed, and steps towards reform, if they are to have meaningful effects, must be considered and evaluated in a disciplined, rigorous, and initially skeptical way. We need greater trans-

4 Stacy L. Hawkins, *Mismatched or Counted Out? What's Missing from Mismatch Theory and Why it Matters*, 17 U. PA. J. CONST. L. 863 (2015).

5 For example, see the Hawkins and Sander discussion of Questions 11 and 12 in Arcidiacono et al., *supra* note 2, at 721–28.

6 I coauthored what is still perhaps the most definitive study of law school academic support programs and pioneered a system of class-based affirmative action to foster greater socioeconomic diversity at my law school. See Kristine S. Knaplund & Richard H. Sander, *The Art and Science of Academic Support*, 45 J. LEGAL EDUC. 157 (1995); Richard H. Sander, *Experimenting with Class-Based Affirmative Action*, 47 J. LEGAL EDUC. 472 (1997).

parency, and we need to zealously evaluate reforms, or we will find ourselves simply substituting a new, soft ideology for the old one, with real progress in reducing mismatch, and improving minority student outcomes, as far away as ever.

I. CLARIFICATIONS AND CONUNDRUMS

Although Professor Hawkins often tries to be fair to mismatch theory, at other points she seriously mischaracterizes it, probably more out of misunderstanding than malice. To the extent that she is reflecting common misunderstandings of mismatch, I welcome this opportunity to clarify the record.

A. *The Naiveté Argument*

Hawkins suggests that mismatch theory suffers from “tunnel vision,”⁷ by focusing too much on the relationship between academic credentials and higher education outcomes. “Mismatch theory is alluring in its simplicity . . . [but the] theory’s causal claims linking academic credentials to academic performance actually rely on a series of intermediate assumptions.”⁸ “Mismatch theory’s isolation from the rich store of data and research in the fields of cognitive and developmental psychology is problematic given the intersections between this research and the claims on which mismatch theory is based.”⁹

These statements are wrong both in letter and in spirit. Some of the earliest social scientists to empirically test the mismatch hypothesis *were* cognitive psychologists, such as Rogers Elliot at Dartmouth University and Frederick Smyth at the University of Virginia.¹⁰ These scholars found that students who had academic credentials (measured in a variety of ways, not just with test scores) far below those of their classmates, tended to drop out of science majors at rates that were not simply higher than those of their classmates with stronger credentials, but higher than the rate for similar “comparison” students at other institutions who did not have a large credential gap with their classmates.¹¹ These psychologists did not start out intend-

⁷ Hawkins, *supra* note 4, at 861.

⁸ *Id.*

⁹ Hawkins, *supra* note 4, at 857.

¹⁰ See Rogers Elliott, A. Christopher Strenta, Russell Adair, Michael Matier & Jannah Scott, *The Role of Ethnicity in Choosing and Leaving Science in Highly Selective Institutions*, 37 RES. HIGHER EDUC. 681 (1996); Smyth & McArdle, *supra* note 1.

¹¹ Elliott, Strenta, Adair, Matier & Scott, *supra* note 10, at 701 tbl. 2; Smyth & McArdle, *supra* note 1, at 369 fig. 4.

ing to make an argument about academic credential gaps and mismatch; nor were mismatch hypotheses the only ones they tested. Rather, the idea of mismatch (in this case, “competition” mismatch) *found them*—it arose from their careful examination of the data and the models.

Similarly, the remarkable randomized experiments at the Air Force Academy¹² (described in Part IV, below) were not aimed at studying mismatch, but at testing an innovative approach to peer learning. The experimental results, perplexing at first, led the researchers to investigate mismatch (in this case “social mismatch”) as a possible explanation, which indeed robustly explained the results. This is hardly “tunnel vision”; it is, rather, scientific inquiry at its best.

Yet another example is the experimental research of Esther Duflo and her colleagues. A very large, randomized experiment showed that learning decreased as the measured academic preparation level of students in a classroom increased in heterogeneity.¹³ There are no “intermediate assumptions” in the Duflo study, which was published in the *American Economics Review*. It is a scientific, randomized experiment that admits of no alternative explanation.

Professor Hawkins never shows that any of the mismatch research is incorrect, or that the results presented in any mismatch paper would somehow change if different explanatory factors were inserted into the regression equations. Thus, to the extent her argument is that mismatch *findings* are wrong or incomplete, such assertions are unfounded and unsupported. I suspect Hawkins is confusing the “positive” and “normative” dimensions of the affirmative action debate. As a “positive” matter, the first-order effects of mismatch have been rigorously established with scientific studies; serious consequences follow when we create large credential gaps among students. Normatively, there are many ways we might address mismatch problems (and other performance gaps across different groups of students), and I enthusiastically agree with many of Hawkins’ reform proposals, as long as it is understood that some kind of change is needed in the status quo, and the reform options we take seriously

12 Scott E. Carrell, Bruce I. Sacerdote & James E. West, *From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation*, 81 *ECONOMETRICA* 855 (2013). See also notes 62–66 and accompanying text.

13 Esther Duflo, Pascaline Dupas & Michael Kramer, *Peer Effects, Teacher Incentives, the Impact of Tracking: Evidence from a Randomized Experiment in Kenya*, 101 *AM. ECON. REV.* 1739 (2011).

should be implemented on an experimental basis and rigorously evaluated.¹⁴

B. Race and Mismatch

A major theme of Professor Hawkins' article is that much of the underperformance attributed to mismatch is actually due to environmental conditions on campuses affecting minority (especially black) students, which she mostly groups into "stereotype threat" and "stigma threat" issues.¹⁵ This argument is based on a fundamental misunderstanding of the mismatch literature. In the bulk of this literature, and in analyses dealing with all three types of "first-order" mismatch effects, there are no differences between blacks, whites, and Hispanics in the way that mismatch operates.¹⁶ Or, to put this more precisely, "race" tends to be a non-significant predictor of performance once mismatch effects are held constant.

This should be seen as good news, since racial problems are often so hard to diagnose and effectively address. Indeed, one of the reasons that I wrote my original paper on law school mismatch was to demonstrate that there was nothing peculiar to African-American law students—aside from the fact that they tended to receive very large admissions preferences—that was causing them to fail the bar at higher rates than comparable white students.¹⁷ But it is also very powerful evidence against the claim that stereotype and stigma threat, to the extent they exist at all in real world settings, are having much impact upon the academic performance of minority students. If learning mismatch, competition mismatch, and social mismatch all operate at a similar level across racial lines—as many studies have now shown they do—then this crowds out any explanatory power for stereotype threat and stigma threat.

¹⁴ Hawkins often implies that mismatch scholarship is brimming over with normative arguments for eliminating or curtailing racial preferences. But the vast majority of this scholarship (including, for example, the leading works cited at note 1 *supra*) either make no normative arguments at all, or advance very mild suggestions, such as providing greater transparency to students applying to college.

¹⁵ Hawkins, *supra* note 4, at 873–83.

¹⁶ This is the finding of many mismatch theorists and most mismatch studies that have specifically considered the question. See, e.g., Smyth & McArdle, *supra* note 1; see also Peter Arcidiacono, Esteban M. Aucejo & Ken Spenner, *What Happens After Enrollment? An Analysis of the Time Path of Racial Differences in GPA and Major Choice*, 1 IZA J. LABOR ECON. 5 (2012), available at <http://www.izajole.com/content/1/1/5>; Richard H. Sander, *A Systemic Analysis of Affirmative Action at American Law Schools*, 57 STAN. L. REV. 367 (2004).

¹⁷ Sander, *supra* note 16, at 429, 449–54 (discussing the author's motivation behind writing the article).

C. *First-order and Second-order Mismatch Effects*

As explained in footnote 2 above, and at greater length elsewhere,¹⁸ I and other mismatch scholars have found it quite useful to distinguish “first-order” mismatch effects—things that result directly from credential gaps among students—from “second-order” effects, which are indirect consequences of the first-order effects. “Learning mismatch” is an example of a first-order effect; students learn less in an environment where they have levels of academic preparation far below most of their peers. Depressed graduation rates are a second-order effect; a student might be less likely to graduate from college if they are learning less, but, then again, they might graduate anyway if the learning mismatch is not very severe, or if the college tries hard to graduate every student. The point of distinguishing among different types of first-order effects (learning mismatch, competition mismatch, and social mismatch) and the more contingent second-order effects is to make more precise just what we are talking about when we discuss mismatch, and also to make clear that first-order effects, because they are direct effects, will tend to show up more consistently in the empirical literature.

Hawkins advances a critique that seems linked to this first-order/second-order distinction. She writes,

according to mismatch theory, weaker relative academic credentials generate the following intermediate forms of mismatch: learning mismatch, competition mismatch, and social mismatch. It is these three intermediate forms of mismatch that ultimately [cause] academic underperformance. But it is important to note that it is academic underperformance, and not these intermediate forms of mismatch, which is the crux of the harm claimed to accrue to URM students under mismatch theory.¹⁹

These statements are wrong in several different ways. The first-order effects are not “intermediate”; they *are* the actual mismatch phenomena. They cause direct harm to students. Whether they produce academic underperformance depends on circumstances. For example, social mismatch directly reduces social interaction between students at the same college who have very different levels of academic preparation. This can cause students to suffer academically,²⁰ but the direct and primary harm is a reduction in interaction—which, when it correlates with race, reduces the benefits of racial diversity on campus.²¹

¹⁸ See Sander, *supra* note 2, at 1642–43.

¹⁹ Hawkins, *supra* note 4, at 861–62.

²⁰ See *infra* Part IV.

²¹ See Peter Arcidiacono, Esteban Aucejo, Andrew Hussey & Ken Spenner, *Racial Segregation Patterns in Selective Universities*, 56 J. LAW AND ECON. 1039, 1040 (2013) (“The benefits

Or consider competition mismatch, which often causes students with low relative academic preparation to abandon a preferred field (often in the sciences or engineering) for one less competitive (such as communications).²² It does not follow that the affected student will academically “underperform” in terms of, for example, GPA or graduation outcomes—indeed, the student who switches to communications may have a higher GPA and an easier time graduating. But there is still a harm if mismatch means the student must abandon her dream of becoming, say, an engineer.

II. IMPROVED MEASURES OF ACADEMIC POTENTIAL

A central tenet of Professor Hawkins’ argument is that mismatch critics are overly focused on the standardized test score gap between racial minorities and other students (for simplicity of empiricism and exposition, I will focus on “African-American/black” and “Anglo/white” students).²³ Hawkins believes that these gaps overstate the difference in academic potential between blacks and whites, and that the gaps also overstate the degree of preference that universities use in admitting African Americans.²⁴ She also believes that by focusing on aspects of student merit, such as “grit,” one can identify students whose abilities greatly exceed whatever might be indicated by their test scores, and thus neutralize the test-score gap.²⁵

Here as elsewhere, I thoroughly agree with part of Hawkins’ claim: we most certainly *should* devote effort and resources towards finding and developing alternate measures of student potential.²⁶ The current metrics used by colleges and graduate schools are only moderately successful in predicting success (though they are a good deal

derived from student diversity, however, will likely not merely depend on the racial composition of the student body but also on the frequency and intensity of social interaction and friendship among students of different races.”); Peter Arcidiacono, Shakeeb Khan, & Jacob L. Vigdor, *Representation versus Assimilation: How do Preferences in College Admissions Affect Social Interactions?*, 95 J. PUB. ECON. 1, 1 (2011) (discussing the actual existence of diversity on campuses).

²² See Smyth & McArdle, *supra* note 1, at 354–57 (citing studies that reached this conclusion).

²³ *E.g.*, Hawkins, *supra* note 4, at 862–63.

²⁴ Hawkins, *supra* note 4, at 863.

²⁵ Hawkins, *supra* note 4, at 868–70.

²⁶ I did something similar at UCLA Law School by creating a measure of college GPA that took into account grade inflation and competitiveness in college; this so-called “national grade” gave less weight to highly elite schools (such as Stanford and the Ivy League) than did UCLA’s former measures (and thus helped low-SES applicants, on balance). It also did a substantially better job of predicting law school performance.

better than is often claimed);²⁷ it is clear we could do at least somewhat better. And the current metrics tend to be skewed in ways that favor more affluent students (though they are less skewed than the college selection methods they replaced).²⁸ Better measures deserve our investment and support.

At times, Hawkins seems to imply that colleges and universities are already using these improved measures in assessing students, and that the apparent racial disparities suggested by test-score differences are therefore much smaller in practice.²⁹ In other words, she suggests that current affirmative action programs do not involve very large preferences, but smaller preferences based on the insight of admissions officers into the more subtle achievements of black and other minority applicants.³⁰ However, Hawkins offers no evidence on this point, and there is, in fact, not the slightest *quantitative* evidence that this is true. On the contrary, where data is available on the admissions decisions of selective universities, test scores, high school grades, and achievement tests clearly account for the bulk of admis-

27 For example, it is often pointed out that entering credentials (for example, some combination of test scores and pre-entry grades) predict only about 20% of the variation in student performance, and sometimes even less. But such a statistic ignores the “restriction in range” problem: students within an academic program, since they are all selected by these credentials, have very little variation in their test scores and prior grades—and these are thus necessarily weak predictors of performance at school. In situations that avoid the restriction of range problem, “credentials” do a much better job of prediction. Sander, *supra* note 16, at 420. An interesting example comes from large law firms; law school grades strongly predict which first-year associates eventually make partner, though law firm partners clearly do not consider law school transcripts when they award new partnerships. See Richard Sander & Jane Bambauer, *The Secret of My Success: How Status, Eliteness, and School Performance Shape Legal Careers*, 9 J. EMPIRICAL LEGAL STUD. 893, 911 (2012) (“GPA is strongly predictive of one’s survival at the big firm.”).

28 Nicholas Lemann’s *The Big Test: The Secret History of American Meritocracy* (1999), explores how the SAT emerged in significant part out of efforts by elite colleges to develop a more “meritocratic” approach to admissions that would expand opportunities for students outside private schools and other traditional feeder schools for the Ivy League, such as working-class students, children of immigrants, and those in rural areas. And while there are large “class” disparities in performance in standardized tests, most of this appears to accurately reflect differences in academic preparation. There is little evidence that low-SES students on average outperform their college (or graduate school) performance as predicted by standardized test scores and grades. See THOMAS J. ESPENSHADE & ALEXANDRIA W. RADFORD, NO LONGER SEPARATE, NOT YET EQUAL: RACE AND CLASS IN ELITE COLLEGE ADMISSION AND CAMPUS LIFE, 250–51 (2009) (reporting results of a large-scale analysis of students at selective colleges and finding that social class does not significantly predict higher or lower class rank).

29 Hawkins, *supra* note 4, at 865–67.

30 *Id.*

sions decisions—and this is, if anything, generally more true of black applicants than white ones.³¹

Moreover, when we predict the performance of college or graduate students based on their standard entering credentials (test scores and earlier grades), we find that black and Hispanic students either perform at the level predicted by their credentials, or somewhat underperform.³² (Hawkins suggests this underperformance reflects the effects of stereotype or stigma threat; I discuss elsewhere in this Article why this does not seem to be true, but for now let us set it aside.³³) If colleges were using other indicia of academic promise that tended to correct for the racial test-score gap, then it follows that black and Hispanic students would *outperform* those traditional credentials.

The one example of which I am aware in which higher education leaders were able to improve racial representation *and* reduce racial performance gaps, is at the University of California. Some time ago, the university began to calculate an “adjusted” measure of high school grades (“UC-adjusted HSGPA”), which considered a standard set of courses from California high schools, and made uniform adjustments to GPAs based on the relative difficulties of the courses (giving more weight to “honors” and “AP” classes).³⁴ Purged of much of the noise that plagues high school GPA measures, the UC-adjusted HSGPA was a far better predictor of college grades than the unadjusted measure, and better even than SAT or ACT scores (though using *both* test scores and UC-adjusted HSGPA produced still better predictions).³⁵ Moreover, using this measure disproportionately benefited black and Hispanic applicants to UC schools.³⁶ For all of these reasons, this innovation is an important example of what universities can and should do to achieve the kind of goal Hawkins advocates.

³¹ The best available data on this point comes from law schools. My analysis of admissions data from forty public law schools over admissions cycles between 2005 and 2007 found that a combination of LSAT scores, undergraduate grades, and race could account for 86% of all admissions decisions. If one examined only black applicants, then LSAT scores and undergraduate grades *alone* accounted for an average of 88% of all admissions decisions. See Richard H. Sander, *Why Strict Scrutiny Requires Transparency: The Practical Effects of Bakke, Gratz, and Grutter*, in *NEW DIRECTIONS IN JUDICIAL POLITICS* 296 (Kevin McGuire ed., 2012) (discussing the fact that schools wanted to lean heavily on test scores and grades).

³² This is shown in a wide array of studies, including those cited at notes 55–56 *infra*.

³³ See *infra* text accompanying notes 48–60.

³⁴ See Kate Antonovics & Ben Backes, *The Effect of Banning Affirmative Action on College Admissions Policies and Student Quality*, 49 *J. HUM. RESOURCES* 295, 296 (2014) (examining “the end of race-based affirmative action at the University of California”).

³⁵ Calculation by the author based on the 2008–10 UCOP database (available on request from the author).

³⁶ Antonovics & Backes, *supra* note 34.

But it should also sober our expectations, for using the UC-adjusted HSGPA only offset between 3% and 10% of the effects of not using racial preferences at UCLA and Berkeley after Proposition 209.³⁷

The broader notion that racial disparities simply reflect inadequacies in how we measure academic preparation is not well supported. The National Assessment of Educational Progress (“NAEP”) performs very extensive testing to capture trends in learning in American schools; unlike standardized tests such as the SAT, NAEP’s measures are based on subject-matter learning in specific fields covered by virtually all schools. I have discussed these results at greater length elsewhere,³⁸ but the simple bottom line is that the NAEP-measured racial gaps closely approximate those captured by the SAT.³⁹ The median African-American in twelfth grade is at a level of academic preparation comparable to the median white eighth grader.⁴⁰

One of the most significant efforts ever undertaken to reduce racial bias in the measurement of learning was the development of the “performance test” component of the California Bar Exam in the early 1980s.⁴¹ Research consistently showed a large racial disparity in performance on traditional bar exams, which consisted of a “multi-state” multiple choice exam and a series of “essay” questions analogous to law school final exams. Both were subject to the criticisms that they measured “academic” rather than “practical” knowledge, and were likely to disadvantage minority students who might do poorly on standardized tests but would actually make fine lawyers. The California Bar set about experimenting with a variety of other evaluation strategies that might address these concerns, and through very extensive research came up with the “performance exam.” In this one-day test, bar-takers would open a file of materials that simulated

³⁷ *Id.* at 310.

³⁸ See RICHARD SANDER & STUART J. TAYLOR, MISMATCH: HOW AFFIRMATIVE ACTION HURTS STUDENTS IT’S SUPPOSED TO HELP, AND WHY UNIVERSITIES WON’T ADMIT IT 261–62 (2012).

³⁹ See NAT’L CTR. FOR EDUC. STATISTICS, NATIONAL ASSESSMENT OF EDUCATION PROGRESS 2008: TRENDS IN ACADEMIC PROGRESS 15 (2009).

⁴⁰ NAT’L CTR. FOR EDUC. STATISTICS, NCES 2013-456, NAEP 2012 TRENDS IN ACADEMIC PROGRESS (2013). The National Assessment of Educational Progress (“NAEP”) is the nation’s largest effort to develop systematic measures of learning in the nation’s schools and make those measures comparable over time. According to this quadrennial overview report, the median reading score for 13-year-old whites was 270 in 2012; the median reading score for 17-year-old blacks was 269. *Id.* at 17–18. The median NAEP mathematics score for 13-year-old whites was 293 in 2012; the median mathematics score for 17-year-old blacks was 288. *Id.* at 39–40.

⁴¹ This story is described in Stephen P. Klein’s *An Analysis of the Relationship Between Trial Practice Skills and Bar Examination Results* 1–10 (1983).

the sort of real-world tasks lawyers regularly undertake: a memo outlining a client's problem and a variety of other evidentiary and legal materials that the test-taker would use to fashion a strategy for her client.

This performance test, which was permanently incorporated into the California Bar exam and has gradually been adopted by other jurisdictions, did succeed in making the bar exam more relevant to actual lawyer problems; it may also have given law schools a reward for and incentive in developing clinical programs. But it did nothing to reduce the racial performance gap, which was exactly the same size on the performance exam as on the other portions of the bar exam.⁴²

Another important real-world experiment reinforces this finding. In the early 2000s, UCLA initiated an admissions process through which readers evaluated applications on three separate dimensions: academic achievement, life challenges, and "personal achievement"—this last category focusing on such things as contributions to one's community, extra-curricular activities, and talents and skills not captured by ordinary academic measures.⁴³ The implication of the Hawkins critique is that with such a measure, underrepresented minorities with so-so academic credentials would have higher-than-average levels of personal achievement (so that an admissions process taking those into account would correct racial disparities). Certainly UCLA, which was trying hard to overcome the effects of a ban on the use of racial preferences, had every incentive to produce such results. But as it turns out, race did not predict levels of personal achievement.⁴⁴ What did predict personal achievement was, of all things, academic achievement; students who had academic achievement levels a standard deviation above the average of UCLA applicants had personal achievement levels that were a third of a standard deviation above the average.⁴⁵

Professor Hawkins mentions some valuable initiatives that certainly deserve further investigation, experimentation, and support. The Posse program, as she suggests, has been successful in identifying minority students with high levels of "grit," and this does indeed predict success in college.⁴⁶ But the percentage of students so identified is so

⁴² *Id.* at 12.

⁴³ Richard H. Sander & Medha Uppala, Racial Discrimination in UCLA Undergraduate Admissions: An Inquiry in Depth 10–11 (Sept. 15, 2014) (unpublished manuscript) (on file with author).

⁴⁴ *See id.* at 11.

⁴⁵ *Id.* Further analysis has since been conducted by the author.

⁴⁶ *See* Hawkins, *supra* note 4, at 889–92.

small that there are real questions about how widely this model can be generalized. What would be the cost, and the feasible payoff, of institutionalizing this on a scale large enough to replace a substantial part of current preferential admissions? Similarly, Hawkins cites approvingly the work done by scholars at Berkeley on the usually-overlooked characteristics of lawyers that predict success in their careers.⁴⁷ This work, too, is interesting and worthy of further support and (in particular) large-scale experimentation. But to date, the Berkeley project has not produced a workable alternative to current law school admissions systems, partly because the main purpose of current law school admissions procedures is to identify students academically able to do the work, and the subjects of the Berkeley research have all already met those criteria. The Berkeley studies need to experiment with actual law school admissions, not only to see to what extent applicants can “game” the more subjective criteria used, but also to determine whether these criteria can actually identify students who will do well in law school and whether the criteria produce a significant racial dividend.

III. STEREOTYPE THREAT AND STIGMA THREAT

A second pillar of Professor Hawkins’ argument is that underrepresented students often perform below their potential because of “stereotype” threat and “stigma” threat, both of which undermine students’ self-confidence and make it more difficult for them to perform at their best. If we can counter these by reducing prejudice and hostility among “majority” students on campus, then the initiatives aimed at picking the best minority students will lead to very strong performance, and eventual college and career success.

There is certainly a good deal of intriguing *laboratory* evidence that stereotype threat in particular can undermine performance of all sorts of groups (the evidence is much weaker for stigma threat).⁴⁸ But there are four enormous obstacles to taking these phenomena seriously as part of the type of strategy Hawkins proposes.

First, as I noted in Part I, stereotype threat and stigma threat cannot explain mismatch because mismatch is not, in significant measure, a racial phenomena. It disproportionately affects some minorities because of the operation of racial preferences in admissions, but

⁴⁷ Hawkins, *supra* note 4, at 906.

⁴⁸ An early, classic analysis is Claude M. Steele & Joshua Aronson, *Stereotype Threat and the Test Performance of Academically Successful African Americans*, in *THE BLACK-WHITE TEST SCORE GAP 401* (Christopher Jencks & Meredith Phillips, eds., 1998).

mismatch studies have shown again and again that mismatch affects whites who receive preferences too,⁴⁹ and that mismatch does not affect blacks who do not receive preferences, or who receive smaller preferences.⁵⁰

Second, there is a good deal of evidence that stereotype threat does not have large effects in real-world settings. Paul Sackett, whose work is cited approvingly by Hawkins in another context,⁵¹ has undertaken with colleagues several experiments aimed at creating a degree of “stereotype threat” in real-world testing situations.⁵² His results strongly suggested an absence of stereotype threat. My own research on legal education finds that the black/white gaps in performance are the same (or higher) in evaluation settings with low stereotype threat (i.e., legal writing classes where students work in loose collaboration with a small section instructor) and those with high stereotype threat (timed, all-important, end-of-semester final exams).⁵³ There are good conceptual reasons to explain the lack of translation of these results from the lab to the real world.⁵⁴

Third, the evidence that minority undergraduates “underperform” their credentials in college is not particularly good evidence that these students’ grades are undermined by stereotype threat and

49 See, e.g., Smyth & McArdle, *supra* note 1, at 369 (showing virtually identical patterns for whites and underrepresented minorities); Peter Arcidiacono, Esteban Aucejo & Ken Spenner, *What Happens After Enrollment? An Analysis of the Time Path of Racial Differences in GPA and Major Choice*, 1 IZA J. LAB. ECON. 5 (2012) (noting that although blacks at Duke drop out of science and economics majors at much higher rates than whites, “accounting for academic background can fully account for differences in switching behaviors across blacks and whites”).

50 See *supra* note 16 and accompanying text. Hawkins responds to this point by suggesting that the result of “good outcomes” for minorities in low-preference environments is driven by blacks at HCBUs, where the better outcomes do not reflect the absence of preferences, but the absence of racism. Hawkins, *supra* note 4, at 878–80. This is a thoughtful point, but a close look at the literature shows that it does not hold. Blacks in lower- or low-preference environments at majority-white institutions have the same amount of improvement in performance outcomes that the mismatch hypotheses predict. Doug Williams, for example, has run his “first-choice/second-choice” analyses with HCBU law schools excluded from the analysis, and finds exactly the same benefits to blacks attending a second-choice school using lower preferences (the standard errors in his analysis go up, reflecting the smaller sample size, but the coefficients remain identical).

51 See Hawkins, *supra* note 4, at 866 n.34.

52 See Paul R. Sackett & Ann Marie Ryan, *Concerns About Generalizing Stereotype Threat Research Findings to Operational High-Stakes Testing*, in STEREOTYPE THREAT: THEORY, PROCESS, AND APPLICATION 249 (Michael Inzlicht & Toni Schmader eds., 2012).

53 See Sander, *supra* note 16, at 424.

54 Reasons for the lack of translation are provided in Sackett & Ryan, *supra* note 52, and include the likelihood that stereotype “triggers” are noisier in real-world settings, and students are putting forth more serious and uniform effort on tests that actually affect their futures.

stigma threat. Rather, it suggests poor measurement of their credentials. The studies that have found the strongest evidence of minority underperformance, such as William G. Bowen and Derek Bok's *Shape of the River*, tend to have very weak controls for both the quality of a subject's high school and the difficulty of a student's curriculum.⁵⁵ If black college freshmen (to take the simplest example) have attended substantially weaker high schools and have generally taken less rigorous courses than their white classmates (both of which we know to be true), then of course a generic measure of credentials like unadjusted high school GPA will tend to overpredict black performance. When we look instead at law students, we find much lower levels of black underperformance—so low that the differences with whites are often not even statistically significant.⁵⁶ The explanation? The black and white applicants to law school are coming from a largely overlapping set of institutions, and their curricula, though still somewhat divergent, are much more similar across racial lines than is the case for high school. Thus, the significantly better comparability of college GPAs across racial groups produces better predictions and thus a smaller “underperformance” gap.

But the clincher in this argument comes with bar exams. Repeated studies that control for LSAT scores, first-year law school GPAs, and the actual law school that students have attended find *no evidence* of black underperformance.⁵⁷ Race, so far as psychometricians can tell, is completely irrelevant to bar scores once we take into account

55 See, e.g., WILLIAM G. BOWEN & DEREK BOK, *THE SHAPE OF THE RIVER: LONG-TERM CONSEQUENCES OF CONSIDERING RACE IN COLLEGE AND UNIVERSITY ADMISSIONS* 72–78 (1998) (finding strong evidence of minority underperformance, but relying only on very broad measures of credentials, such as grouping high school grades into three broad categories and ignoring measures such as high school quality). Both of these factors will significantly bias upward predictions, based on credentials, of African-American students' performance.

56 A comprehensive analysis by LSAC in 2003 found that black “underperformance” was about one-eighth of a standard deviation in first-year grades, or about one-sixteenth of the observed gap in black/white performance. LISA C. ANTHONY & MEI LIU, LAW SCH. ADMISSION COUNCIL, *ANALYSIS OF DIFFERENTIAL PREDICTION OF LAW SCHOOL PERFORMANCE BY RACIAL/ETHNIC SUBGROUPS BASED ON THE 1996–1998 ENTERING LAW SCHOOL CLASSES*, LAW SCHOOL ADMISSION COUNCIL LSAT TECHNICAL REPORT 00–02, at 9 (2003). In my own subsequent analyses, I found that adding measures of college quality and grade inflation eliminated any black/white gap. See Richard H. Sander, *A Reply to Critics*, 57 *STAN. L. REV.* 1963, 1968 (2005).

57 This was the conclusion of the enormous Bar Passage Study undertaken by LSAC in the 1990s. See LINDA F. WIGHTMAN, LAW SCH. ADMISSION COUNCIL, *LSAC NATIONAL LONGITUDINAL BAR PASSAGE STUDY* 80 (1998). This has been confirmed by a wide variety of independent research. See, e.g., Stephen P. Klein, *Disparities in Bar Exam Passing Rates Among Racial/Ethnic Groups: Their Size, Source, and Implications*, 16 *T. MARSHALL L. REV.* 517, 529 (1991); Sander, *supra* note 16, at 429.

pre-law credentials and law school grades. Yet bar exams, above nearly any other academic setting, should be very ripe for real-world stereotype threats, if they exist: blacks sitting for the bar know very well that black bar passage rates are dramatically below those of whites.⁵⁸ Why does race instead have no effect at all upon bar scores? Because, consistent with the college and law school stories told above, the bar analyses have the most accurate measures of pre-bar credentials: results from standardized LSAT exams, and from essentially identical curricula and grading systems at identical law schools. In other words, the better our controls for real academic preparation levels, the more race appears to disappear from the performance equation. Until convincing reasons can be offered for why these heavily-documented patterns produce results completely inconsistent with the “threat” theories, we should not rely on such theories as central guides to our admissions or education policies.

Finally, it is vital to keep in mind that the policies Hawkins and others advocate based on “stereotype” threat and “stigma” threat have an immensely counterintuitive problem at their cores. If these threats are real and truly undermine minority performance, the obvious policy implication is that on college campuses, administrators should bend over backwards to avoid using disparate admissions standards across racial lines. For what could more robustly activate these threats than using very different admissions standards based on race?⁵⁹ What could more surely produce disparities in performance that nourish stereotype threat?⁶⁰ What could arouse the type of focus

58 WIGHTMAN, *supra* note 57, at 80 (presenting the author’s analysis of LSAC data from the Bar Passage Study, which found much greater levels of concern about passing the bar among black respondents than among whites).

59 See JAMES SIDANIUS ET. AL., THE DIVERSITY CHALLENGE: SOCIAL IDENTITY AND INTERGROUP RELATIONS ON THE COLLEGE CAMPUS 317 (2008) (finding that in certain circumstances students who believed they had been admitted due to racial preferences were at significantly higher risk of stigma from their white classmates).

60 A key piece of evidence that Hawkins relies on to counter this point instead illustrates the shocking shoddiness of much of the research in this area. Angela Onwuachi-Willig, Emily Houh & Mary Campbell, *Cracking the Egg: Which Came First—Stigma or Affirmative Action*, 96 CALIF. L. REV. 1299 (2008), attempts to demonstrate empirically that minority students experience the same type of stigma effects in schools that do not use racial preferences, as in schools that do. But this is “demonstrated” through a sample of seven law schools, including three (UC Berkeley, UC Davis, and University of Michigan) operating in states where racial preferences were illegal. *Id.* at 1305. At these schools, students were invited to take a self-administered web survey, which described itself as a study of stigma experienced by minorities. The response rate at the schools operating in “race-neutral” states was only 27% (and dramatically lower than in the “control” states—an intriguing finding that the authors ignore) and the number of blacks participating at these schools seems to have been a dozen or fewer—all reasons not to take the study seriously. Its greatest problem, however, was that it made no attempt to measure whether the schools

on double standards that activates stigma threat? Yet Hawkins asks us to essentially disregard these overwhelmingly common-sense instincts and assume that these problems can be dealt with by changing campus climate and educating racial generalizations out of the majority student population.

IV. RESPONSIBLE, ACCOUNTABLE SOCIAL ENGINEERING

A recent series of events at the United States Air Force Academy helps to put into strong relief a fundamental difference between Professor Hawkins' approach to these issues and my own.

Three economists (two of them on the faculty at the Academy) undertook a study in the mid-2000s of the effect peers had on cadet learning.⁶¹ They discovered that, within relatively small groups (squadrons of about thirty students), cadets learned more when they were in a squadron whose average academic achievement was somewhat higher than their own. In other words, academic "peer effects" were positively related to the academic strength of one's peers, if the peer group was relatively small. This was an important finding, it made intuitive sense, and it complemented nicely many of the ideas Professor Hawkins has advanced. Mismatch effects might be powerful in large or anonymous classrooms—where the effect of being academically weaker meant that one might be left behind in class—but they could perhaps be offset in structured, small peer groups, where one could benefit from informal study groups or tutoring from academically stronger students.

The economists used their peer-group findings to persuade the Air Force Academy to undertake a large, randomized experiment with incoming cadets—something all too rare in higher education.⁶²

involved actually *were* race neutral—a rather significant problem, since all the available data shows that law schools in states with preference bans, and two of these three law schools in particular, *continue to use* racial preferences. See, e.g., Danny Yagan, *Affirmative Action Bans and Black Admissions Outcomes: Selection-Corrected Estimates from UC Law Schools*, (Univ. of Cal. at Berkeley & Nat'l Bureau of Econ. Research, Working Paper No. 20361, 2014) ("[T]he affirmative action ban far from eliminated cross-sectional black admission advantages . . ."). A significant problem in law school discourse, and one that the Hawkins article perhaps inadvertently contributes to, is the tendency of scholars (not Hawkins) to produce shoddy empirical research, draw far-reaching conclusions from it, publish it in student-edited law reviews (since the research could not withstand serious peer-review), have the research widely cited (here Hawkins is at fault), and thus gain credibility simply through the frequency with which its findings are uncritically repeated.

61 See Scott E. Carrell, Richard L. Fullerton & James E. West, *Does Your Cohort Matter? Measuring Peer Effects in College Achievement*, 27 J. LABOR ECON. 439 (2009).

62 See Carrell, Sacerdote & West, *supra* note 12, at 855 (discussing the parameters of the study).

The experiment was generously supported by blue-ribbon overseers from the National Science Foundation, the National Academy of Education, and the Spencer Foundation. In the experiment, hundreds of cadets were randomly assigned to squadrons, while hundreds of others were assigned to squadrons designed to maximize the academic performance of those students arriving at the Academy with the weakest academic preparation.⁶³ These students were assigned to squadrons with a mix of high-performing students, thus creating a small, squadron-sized environment that matched in many ways the large environment at highly-selective universities.

The results were a surprise. The low-preparation students assigned to the experimental groups had substantially worse academic performance than the low-preparation students in the control groups.⁶⁴ In other words, the academic mismatch effects these students experienced were aggravated by the experimental squadron groups that tried to maximize their small-group interaction with high-performing students. The reason? According to the project designers, the low-performing students “avoided the [students] with whom we intended them to interact and instead formed more homogeneous subgroups.”⁶⁵ That is, social mismatch effects aggravated academic mismatch effects. The experiment was declared a failure and terminated. All of those concerned deserve great credit for not suppressing results that were, no doubt, embarrassing on all counts. Instead, the results were documented in detail and published in *Econometrica*, one of the most prestigious and carefully peer-reviewed journals in all social science.

I give this example not simply to cite another example of well-documented mismatch (a particularly powerful example both because of the large-scale, experimental design and because the results were neither expected nor favored by the investigators). My broader point is well-captured by the study authors in their conclusion: “[w]e conclude that social processes are so rich and complex that one needs a deep understanding of their formation before one can formulate ‘optimal policy.’”⁶⁶ These are words we should take to heart, and which have several clear implications. We should launch new social experiments with a sense of humility; they should be set up as

63 A key motivation for the Academy’s support of the experiment was the desire of “USAFAs senior leadership [to] reduce the academic probation rate, then at roughly 20 percent.” *Id.* at 856 n.4.

64 Carrell, Sacerdote & West, *supra* note 12.

65 *Id.* at 855.

66 *Id.* at 881.

controlled experiments, if at all possible, and in any case should be accompanied by maximum transparency, so that we can observe, measure, and try to understand their effects.

The large-scale social experiment known as affirmative action was begun, and has since been conducted, in ways almost directly opposed to these principles. Selective colleges and graduate programs hastily launched these programs in the late 1960s and early 1970s for political reasons, with little or no attempt to experiment with their approaches or evaluate the results. Colleges and universities tended to mimic one another's preference policies, providing little variation through which one could compare effectiveness. And preference programs have tended to be shrouded in secrecy, both in terms of the size and nature of preferences given, and in terms of the results. Steps towards transparency have only just begun, and even now transparency is rare.

Given that so much of this initial transparency has disclosed widespread indications of serious mismatch effects and an almost criminal neglect of these effects by higher education administrators, the appropriate response is not to brush these findings aside in the hope that more "holistic" admissions and more supportive campuses—which exist more in the imagination of the proponents than at any real-world college—will somehow solve the problems. Nor should we give much credence to researchers who have obvious ideological agendas, rely on anecdotes and psychology experiments rather than real-world controlled social experiments, and usually fail to publish their results in peer-reviewed journals. We need to insist that our efforts going forward proceed on the basis of valid science, careful experimentation, and transparency that helps police our efforts.