

2001

Disparate Outcomes by Design: University Admissions Tests

Jay Rosner

Follow this and additional works at: <http://scholarship.law.berkeley.edu/blrj>

Recommended Citation

Jay Rosner, *Disparate Outcomes by Design: University Admissions Tests*, 12 LA RAZA L.J. 377 (2015).
Available at: <http://scholarship.law.berkeley.edu/blrj/vol12/iss2/9>

Link to publisher version (DOI)

<http://dx.doi.org/https://doi.org/10.15779/Z38W65B>

This Article is brought to you for free and open access by the Law Journals and Related Materials at Berkeley Law Scholarship Repository. It has been accepted for inclusion in Berkeley La Raza Law Journal by an authorized administrator of Berkeley Law Scholarship Repository. For more information, please contact jcera@law.berkeley.edu.

Disparate Outcomes by Design: University Admissions Tests[†]

Jay Rosner

I am currently serving as the Executive Director of The Princeton Review Foundation, administering its national programs since 1995. Before that, I was general counsel to The Princeton Review from 1993 to 1995. Prior to that, I was an Executive Director of The Princeton Review responsible for the northwest region, based in Seattle, from 1987 to 1993. I am a member of the New York, New Jersey, Pennsylvania and Oregon bars.

The Princeton Review is the nation's largest SAT preparation company, and the second-largest test preparation company overall. The Princeton Review Foundation is a 501(c)(3) corporation, established by The Princeton Review in 1987, with the mission of providing test preparation courses, materials and related services to low-income and underrepresented minority (URM) students. URM students usually are considered to be African American, Latino and Native American students.

Most of my work week consists of establishing, designing and monitoring intensive admission test preparation courses provided to URM and/or low-income students. I speak regularly at national and regional educational conferences on this topic, and I have testified several times before legislative committees. My pro bono activities include providing legal advice to students and attorneys who are engaged in disputes with test manufacturers. I have written two articles on admission testing for periodicals, and one article on test company disputes currently posted on the internet. I have not testified as an expert at trial or by deposition in any prior case. I am not being compensated for my work in connection with this matter.

My testimony is based upon my experiences over the last twenty years of regular involvement with testing as an attorney, an SAT and LSAT instructor and a critic of admission tests. In approximately 1980 I served as pro bono counsel to an African American college student who filed suit against the Educational Testing Service (ETS), the world's largest and best-known test producer.

This student's claim arose from his high school counselor advising him not to take a preparation course for the SAT. He subsequently saw several of his classmates take a course and improve their SAT scores, and he later learned that his counselor had just passed on to him the often-repeated ETS position that preparation courses were ineffective. His claim was that ETS misled him as to the effects of coaching on the SAT. This was my first foray into the world of admission testing beyond my having taken the SAT, the LSAT and the bar exam.

Since then I have participated as counsel or consultant in several litigation matters against both ETS and Law Services, the producer of the LSAT. I have read widely on testing issues, and have sought out the most knowledgeable people who could better my understanding of minority students' performance on admission tests. All told, I have had thousands of conversations and interactions, continuously over

[†] Expert report submitted on behalf of Intervening Defendants (Student Intervenors), *Grutter v. Bollinger*, 137 F. Supp. 2d 821 (E.D. Mich. March 27, 2001) (No. 97-75928).

the last 13 years, with students before and after they have taken tests like the SAT and the LSAT. These experiences inform my testimony.

My student advocacy includes testifying before legislative committees in California, Texas and New York regarding standardized admission tests. I am a member of the Advisory Board of the University of Pennsylvania's Department of Academic Support, which targets academic services to minority students. I now supervise the administration of intensive SAT preparation courses provided to nearly 10,000 low-income students in a state-subsidized program in California. I also help administer LSAT and other admission test preparation courses in conjunction with preprofessional advisors on many campuses, including the following historically black campuses: Xavier, Morehouse, Spelman, Howard, Clark Atlanta, Morris Brown, North Carolina Central, Florida A&M, Hampton, Prairie View and Bethune-Cookman.

The data on which I rely comes from information I've gathered in administering SAT and LSAT preparation courses over the years, talking with test-takers and reviewing surveys commissioned by test preparation companies. I have also reviewed and considered data from surveys and studies commissioned by the test manufacturers, as well as other literature in the field.

OPINIONS TO BE EXPRESSED

While the majority of my test preparation experience has been with the SAT, I have had significant experience over the years with the LSAT. I have personally observed that there is a high correlation between an individual's uncoached LSAT score and his/her earlier uncoached SAT score. Professor David S. Mann of the College of Charleston reached a parallel conclusion in an unpublished 1992 study he did of students on his campus. A copy of his study is attached hereto as Appendix B.¹ An ETS study in 1963 by William B. Schrader yielded similar results, and is discussed in *The Reign of ETS*.²

Mr. Nairn concluded that tests like the LSAT and the SAT correlate better with each other than with anything else, which is consistent with my observations over the years.

This correlation epitomizes the unusually similar effects and impacts associated with both the SAT and LSAT. Opinions that I render below about the SAT are virtually always applicable in identical ways to the LSAT, and vice-versa.

I.

WRONG ANSWER KEYS AND "FLAWED" QUESTIONS ON THE LSAT

Individual questions chosen by the test manufacturer to appear on the LSAT are picked because they generate desired statistical outcomes. Sometimes these questions are flawed, and sometimes the answers chosen by Law Services are

1. The exhibit, David S. Mann & Thomas S. Gibson, Predicting LSAT Scores from SAT Scores: A Preliminary and Experimental Report (1992) (unpublished), is not included in this publication, but is on file with the author.

2. Allan Nairn & Associates, *The Reign of ETS: The Corporation That Makes Up Minds* 234-235 (Ralph Nader 1980).

incorrect. As an attorney, I represented Chris Laucks, who successfully challenged the viability of an answer to a one question on the February, 1981 administration of the LSAT.

Laucks' situation serves as a good introduction to LSAT questions and their characteristics. Laucks' answer choice was correct, and his answer was the only correct answer to the LSAT question in dispute; the LSAT answer key was flat-out wrong. Law Services admitted their mistake and reviewed the scores of several thousand test-takers, rescoring many of...those tests. In doing the rescoring, Law Services simply omitted the particular question from their calculations because, they asserted, it was "flawed."

It was not flawed, but it was embarrassing for Law Services: the question along with its wrong answer had met LSAT specifications. For example, high-scoring students tended to give the Law Service's credited (wrong) response, while lower-scoring students picked other answers. Criteria for LSAT question selection can be met independent of truth and accuracy, as Mr. Laucks and I both learned.

I also advised Daniel Hinerfeld, who successfully challenged an answer to a question on the June, 1988 LSAT. Again, Law Services had to rescore many tests. Here, unlike the Laucks' case, there was some ambiguity, and two answers were both deemed to be "correct." Human beings construct the LSAT, and human beings make mistakes.

II.

TEST CONSTRUCTION PRINCIPLES: DISPARATE OUTCOMES BY DESIGN

More important than "flawed" LSAT questions or mistakes in LSAT answer keys are the methods by which questions are chosen to appear on the LSAT. All LSAT questions are pretested, so the statistical outcomes of having a particular question appear on the LSAT are well-known in advance: for example, for each question, the test-maker can accurately for[e]tell the total percentage of test-takers who will chose the correct answer. Moreover, the test-maker can predict the percentage of women, African Americans, Latinos, etc., who will choose the correct answer.

The actual task that Law Services performs, year-in and year-out, is accumulating a test full of individually-chosen LSAT questions with foreseeable cumulative effects, which are that, on average:

- a) whites will score higher than blacks;
- b) men will score higher than women; and,
- c) wealthy students will score higher than poor students.

This occurs not by chance; on the contrary, it arises from the fact that virtually all of the individual questions chosen to appear on the LSAT have, in pretesting, "favored" whites, and men, and the wealthy.

III.

CHOOSING INDIVIDUAL TEST QUESTIONS: DISPARATE OUTCOMES BY DESIGN

I cannot give specific examples of the process referred to above by citing LSAT questions, since I am not aware of any rejected LSAT questions that have ever

been released; however, I can give specific examples of the process using SAT questions, and the procedures used are parallel.

Below are two prospective SAT questions that were *eliminated* by ETS from use on the SAT after pretesting because they favored African American students. The first is a sentence completion question, wherein the test-taker must pick the best answer choices to fill in the blanks. The second question requires the test-taker to choose the best antonym.

“(Question:) The actor’s bearing on stage seemed _____ ; her movements were natural and her technique _____ .

(Answers:)

- a. unremitting ... blasé
- b. fluid ... tentative
- c. unstudied ... uncontrived
- d. eclectic ... uniform
- e. grandiose ... controlled

Correct answer: C

Results: 9% more women than men answered this question correctly. 8% more African-Americans than whites answered this question correctly.

Hypothesis: Women and African-Americans tend to do better on questions dealing with the humanities or the arts.”³

“African Americans do significantly better finding the antonym than do whites:

(Question:) RENEGE:

(Answers:)

- a. dispute
- b. acquire
- c. *fulfill**
- d. terminate
- e. relent

*The asterisk indicates the correct answer.

3. Amy Dockser Marcus, *To Spot Bias in SAT Questions, Test Maker Tests the Test*, Wall St. J., Aug. 4, 1999, at B1.

Researche[r]s speculate that African Americans may use the term 'reneege' more frequently than whites because of a pervasive belief that the United States has renegeed on its promises to black Americans."⁴

The test companies use specific procedures to eliminate questions like the two above. One such procedure is called Differential Item Functioning, and it tends to eliminate individual questions that produce what are considered by the test-makers to be significant disparate impacts between racial or ethnic groups. However, no procedure removes questions that more modestly favor one group over another, and selecting a large number of these questions creates a cumulative effect that contributes significantly to the racial score gap.

Why aren't the above questions chosen to appear on the test? If a primary objective were to minimize differences between racial groups taking the test, these questions would be used. If an important objective were to minimize differences between racial or ethnic groups taking the test, more questions like the two above would be written, and test-makers wouldn't rest until their pools of available questions contained many hundreds of these items. As it is, the limited question pools in use are "bleached," and there are insufficient efforts to diversify them.

Just assume, for a moment, that the LSAT test-makers selected for use on the test only those pretested questions that ALL (or nearly all) favored African Americans by 8%, as does the first question above. Those objecting could argue that the LSAT would then be designed to favor African Americans. Yet, that is exactly what happens in the current construction of the LSAT, except that the selection criteria are the reverse: all (or nearly all) of the individual pretested questions selected for use on the LSAT favor whites over African Americans. Disparate results occur not by happenstance, but by design.

"When I use a word,' Humpty Dumpty said, in rather a scornful tone, 'it means just what I choose it to mean – neither more nor less.'

'The question is,' said Alice, 'whether you *can* make words mean so many different things.'

'The question is,' said Humpty Dumpty, 'which is to be master – that's all.'"⁵

LSAT questions are written, developed, edited, pretested and selected for use on the LSAT through a process that foreseeably generates significantly disparate results. These results are rigged against the URM students whom I teach. Unless and until there are substantial modifications to this methodology of LSAT question selection, these foreseeably disparate results will certainly continue.

There was, years ago, one alternative procedure for question selection that mandated the choice of individual test questions (from the available question pools)

4. Pamela Burdman, *Worth of SAT Exam Questioned*, S.F. Chron., Nov. 11, 1997, at A1.

5. Lewis Carroll, *The Annotated Alice: Alice's Adventures in Wonderland & Through the Looking Glass*, at 269 (Martin Gardner ann., Clarkson N. Potter 1960) (1877).

that resulted in the LEAST disparate impacts between racial groups. This was called the "Golden Rule" procedure, set forth in a settlement to litigation begun in the 1970's. The litigation was between the then-manufacturer (ETS) of the Illinois state insurance exam and the Golden Rule Insurance Company.⁶ The parties eventually settled the litigation in 1984. This procedure, to my knowledge, was used successfully only for several years in the construction of the Illinois insurance exam, and, to my knowledge, has not been used on any admission test.

IV. THE FREE THROW ANALOGY

Two years ago I developed an analogy that, I believe, helps those without expertise in testing gain a much better understand[ing] of the role that admission tests play in education. This analogy emerged from several talks on the SAT that I have given over the years to the Division One (College) Men's Basketball Coaches at their annual convention.

Here's the analogy: admission tests like the SAT and LSAT bear the same relationship to academia as free throw shooting has to the game of basketball. The relationships are similar in that both the SAT/LSAT and free throw shooting can appear, to the uninformed, to be closely related to the more complex and comprehensive respective domains of academia and basketball; however, this appearance is very misleading. Those with a more nuanced understanding of these realms know that the skills involved in standardized testing, like those in free throw shooting, are, in fact, significantly different from the skills involved in the more comprehensive domains.

The SAT/LSAT and free throw shooting share numerous characteristics. Both are static, contrived, superficial, quirky, limited and artificial, and exist in a separate time and space from the more comprehensive domains. Both focus on individual performance completely outside of a realistic context of group interaction, with all movement and communication virtually prohibited. Both call into play psychological factors such as confidence in a more dramatic way than the larger domains. Because of their structure and the absence of a group dynamic, both involve highly repetitive tasks taking place in unchanging environments.

The more comprehensive domains of academia and basketball are complex, dynamic, interactive, messy and ever changing, not unlike life itself. To assert that either a 3-digit number on the LSAT or a free throw percentage can meaningfully predict performance in the more complex domains is simply wishful thinking. Professor Claude Steele has pointed out that results on either extreme in testing and in free throw shooting may generally be predictive of performance in the larger domains, but the vast majority of people, by definition, do not have extreme results in testing or in free throw shooting.⁷

The skill sets required to successfully address the SAT/LSAT and those required to address free throw shooting have compelling parallels. Both skill sets are narrow in scope. A significant weakness in either SAT/LSAT performance or free

6. Golden Rule Ins. Co. v. Mathias, 408 N.E.2d 310 (1980).

7. Claude M. Steele, *Thin Ice: "Stereotype Threat" and Black College Students*, The Atlantic Monthly, Aug. 1999, at 54.

throw shooting does not arise within a context in which alternative abilities can be called upon to compensate. A narrow, well-practiced, machine-like approach enhances results in both realms.

By contrast, the skills required to be successful in the larger domains of both academia and basketball are broad, varied, complex and diverse. A weakness in any one skill area usually can be adequately compensated by strength in another. And, neither SAT/LSAT skills nor free throw shooting skills provide direct benefits in the more complex domains, since the tasks in the more complex domains are so different.

No one choosing basketball players would give significant weight to the contrivance represented by their free throw shooting ability, even though this specific skill might become very important toward the end of close games. Yet very significant weight is allocated to the contrivance of the LSAT in competitive law school admissions, even though it is widely accepted that the LSAT only explains, on average, about 16% of the variance in first year law grades.

I contend that free throw shooting, while it often adds a bit of intrigue to the end of a close game, could easily be eliminated from basketball without much impact. Teams could take the ball out of bounds after fouls. In the same way, the LSAT could be eliminated from law school admissions. Admissions committees would be missing a bit of information; however, as other experts such as Professor Claude Steele have pointed out, this LSAT information is of limited value in predicting variance in first year law school grades.⁸ Furthermore, the LSAT generates very significantly disparate scores, on average, along racial and ethnic lines. Intelligent and well-run law school admission committees could make very good decisions without using LSAT scores. Bates, a Maine college with competitive admissions, has been doing SAT-optional admissions successfully on the undergraduate level for over 10 years now, and a number of other competitive colleges such as Franklin & Marshall have followed Bates' lead.

Almost everything I have learned about admission testing in the last thirteen years is consistent with this free throw analogy. I use the analogy often in addressing groups of minority high school students. They agree that one's free throw shooting percentage can improve optimally with highly disciplined practice combined with effective expert instruction, and that the great majority of motivated people are capable of meaningful improvement. The very same factors, in my experience, are applicable in the same way to the SAT/LSAT.

V.

THE SAT/LSAT, TEST PREPARATION AND ALLEGED IMMUTABILITY OF TEST SCORES

My observations in test preparation have indicated that, on average, students' scores on both the SAT and the LSAT can be increased significantly through expert instruction and the extensive practice required in an intensive, high-quality preparation course. Unpublished surveys conducted by Roper Starch for The Princeton Review, and by Price Waterhouse for Kaplan, have reached the same

8. *Id.*

conclusion. A copy of our competitor Kaplan's 1994 Price Waterhouse LSAT survey is attached as Appendix C.⁹

I estimate that the two largest LSAT preparation companies, Kaplan and The Princeton Review, have a combined total of students in their intensive (and expensive) LSAT courses in excess of 25,000 annually. This total represents substantially more than half the number of students who matriculate into law schools every year, even though not everyone who takes these LSAT courses matriculates.

The appended Kaplan survey indicates that the average score improvement for Kaplan LSAT students was 7.2 points, based upon an average initial score of 147.1 and an average final score of 154.3. On a typical LSAT score conversion scale, a 147 is approximately a 33rd percentile score, and a 154 is approximately a 63rd percentile score. The average score gain for this LSAT course studied was approximately 30 percentile points, which most observers would find so large as to be disturbing. The implications for law school admissions are profound. In my experience, score improvements attained in Princeton Review LSAT courses are similar, and have varied only slightly over the years from 1994 to the present.

I have reviewed some of the research that all test manufacturing companies generate regarding test preparation. The producers and marketers of the SAT and the LSAT, as well as those of the MCAT, all have their in-house, self-funded studies that conclude that preparation courses do not produce meaningful score improvements. Certainly it is in the test producers' self-interest to downplay any gains from preparation: if the tests can be "beaten" through preparation, then what are they measuring? And, if scores can be altered significantly in short periods of time, why use them?

Law School admission procedures that give substantial weight to LSAT scores must assume that those scores are reliable and consistent for the individuals submitting scores. My experience in test preparation indicates that there is, on average, significant variability for individuals' scores over short periods of time (i.e., two months) where there is the intervention of intensive preparation with an expert instructor and extensive practice. On average, admission test scores, in my experience, certainly are neither immutable, nor reliable, nor consistent in the presence of effective preparation.

VI.

THE TEST SCORE GAP AND AVOIDANCE

There has been a test score gap on all standardized tests, including admission tests like the SAT and the LSAT, for as long as there have been substantial cohorts of URM students taking the tests.

One factor contributing to the gap that has not been generally well understood is access to high-quality test preparation courses. This is a factor that would affect low-income students as well as underrepresented minority students, because high-quality test preparation is almost always privately provided, and is almost always expensive.

I have observed that underrepresented minority students tend to avail themselves of intensive preparation courses at lower rates than majority students for several reasons, one of which is cost (intensive courses cost \$800-\$1000). Also

9. The exhibit is not included in this publication, but is on file with the author.

relevant is a concept I have proposed and developed which I call avoidance theory. This theory posits that underrepresented minority students avoid not only taking admission tests, but avoid preparing for them.

This avoidance can best be demonstrated from Law Services' statistics which I have reviewed. They show that for the two LSAT administrations given annually in the fall (September/October and December), the majority of white students take the LSAT in the earlier administration and a substantial majority of black students take the later administration. The December administration is the last one accepted by virtually all law schools for admission for the following fall semester, providing its test-takers with no convenient fall-back in case of poor performance, illness, etc. In my opinion, the best explanation for this phenomenon is that black students tend to procrastinate significantly more than whites in order to avoid taking the test.

LSAT preparation courses typically begin six to ten weeks before each test administration. Someone procrastinating in taking the test will likely procrastinate in researching courses, and will often miss the beginning of a course. Avoidance here would diminish the odds that a student would take an intensive preparation course. Avoiding taking the test spills over to avoiding preparation. Finally, given that admission tests are viewed in such a negative light in many minority communities, I've seen this negativity transfer over to test preparation. In summary, antipathy toward standardized testing can provide strong peer and environmental support to minority students for both postponing test-taking until the last possible moment, and for giving short shrift to intensive preparation.

VII. CONCLUSION

From my perspective assisting URM students to improve their admission test scores on the SAT, LSAT and other tests, I observe a system wherein admission tests are actually designed to result in predictably disparate outcomes. I believe that the vast majority of those constructing these tests do not intend this result, but it is nevertheless determined and fixed by the methods of test construction and question selection that they use, and the pools of questions from which they draw.

Test construction that has equity as an objective would use different methods of question generation and selection, would select questions such as the ones set forth above and would make a sustained effort to diversify the question pools that are used.

Compounding the impact of question selection are such factors as test avoidance and differing access to expensive, high-quality test preparation. Add these factors to the many differences between majority and minority students that Professor Claude Steele describes, and the test score gap becomes easier to understand.¹⁰

These factors are all of critical importance only because the LSAT is given such inordinate weight in law school admissions. My colleagues and I can provide LSAT courses only to relatively few URM students, thereby making a very small dent overall in the test score gap. There are at least two systemic solutions to LSAT score disparities: 1) prohibit or severely restrict the use of the LSAT in admissions;

10. Steele, *supra* note 7, at 53,54.

or 2) dramatically increase the LSAT question pools and generate many more questions without disparate impacts, while balancing LSAT question selection so as to prioritize equity.