Book Reviews

Searching for Truth About Testing

The Reign of ETS, The corporation that makes up minds: The Ralph Nader Report on the Educational Testing Service. By Allan Nairn and Associates. *Washington, D.C.*, 1980. Pp. xvii, 553, $30.00.

An Investigation into the Validity and Cultural Bias of the Law School Admission Test, Prepared for the National Institute of Education. By David White. *New York: National Conference of Black Lawyers*, Mar. 31, 1980. Pp. vi, 173.

Law School Admissions Study. By Susan Brown and Eduardo Marenco, Jr. *San Francisco: Mexican American Legal Defense and Education Fund*, 1980. Pp. xi, 90.

David Kaye †

Irma: People always ring the doorbell when I cannot hear it because I am wearing stereo headphones.
Maude: You must be able to hear it; otherwise you could not tell anyone was ringing it.
Maude's response shows that she had assumed that
(A) the doorbell does not ring when Irma is wearing headphones
(B) Irma's visitors never ring the doorbell unless she is wearing headphones
(C) Irma cannot tell that the doorbell has rung unless she has heard it
(D) the doorbell is not functioning properly[1]

What kind of perversity makes us ask prospective law students such questions? Can their answers really tell us anything about

† Professor of Law, Arizona State University. S.B. 1968, Massachusetts Institute of Technology; A.M. 1969, Harvard University; J.D. 1972, Yale Law School.

1. Educational Testing Service (ETS), Law School Admission Test (LSAT), Form 3ALS1, at 1 (1980).

their potential for the study and practice of law? Is it fair to run blacks, Hispanics, or other minority students through such a maze of dialectic? These are among the more important issues raised and treated, with varying success, in three recent studies: *The Reign of ETS*,[2] a Nader-sponsored expose of the Educational Testing Service (*Nader Report*); *An Investigation into the Validity and Cultural Bias of the Law School Admission Test*,[3] a National Conference of Black Lawyers probe into the Law School Admission Test (NCBL Report); and *Law School Admissions Study*,[4] a Mexican American Legal Defense and Education Fund report on law school admissions policies (*MALDEF Report*).

To those old hands in legal education who were present at the inauguration of the Law School Admission Test (LSAT) in 1948, the disquiet reflected in these books must surely inspire a sense of *deja vu*. In truth, the issues are hardly new, and none of the three works, in all of their 845 combined pages, provides much in the way of original data or improved methods for resolving these issues. Yet the concerns that motivate these and other criticisms of standardized testing are enduring,[5] and this most recent wave of criticism comes when Congress, as well as state legislatures, are contemplating proposals to reform standardized testing.[6] Moreover, determining whether a test is biased is a problem not unique to the LSAT. Constitutional or statutory actions that allege discrimination on the part of employers who use quantitatively scored tests are commonplace,[7] and a careful inquiry into whether the LSAT is objectionable can provide insights into the proper use of statistical analysis in such litigation.[8] In this review, therefore, I shall try to assess, in a more or less objective fashion,[9] some of the charges

2. A. NAIRN, THE REIGN OF ETS, THE CORPORATION THAT MAKES UP MINDS: THE RALPH NADER REPORT ON THE EDUCATIONAL TESTING SERVICE (1980) [hereinafter cited without cross-reference as NADER REPORT].

3. D. White, National Conference of Black Lawyers, An Investigation into the Validity and Cultural Bias of the Law School Admission Test, Final Report Prepared for the National Institute of Education (draft Mar. 31, 1980) [hereinafter cited without cross-reference as NCBL Report].

4. S. BROWN & E. MARENCO, LAW SCHOOL ADMISSIONS STUDY (1980) [hereinafter cited without cross-reference as MALDEF REPORT].

5. There is an extensive critical literature on psychological testing that goes back half a century. *See, e.g.*, B. HOFFMAN, THE TYRANNY OF TESTING (1962); P. VERNON, INTELLIGENCE: HEREDITY AND ENVIRONMENT 18-38 (1979).

6. *See* R. BROWN, SEARCHING FOR THE TRUTH ABOUT "TRUTH IN TESTING" LEGISLATION 15-26 (1980).

7. *See* B. SCHLEI & P. GROSSMAN, EMPLOYMENT DISCRIMINATION LAW 65-131 (1976).

8. *See* p. 453 *infra*.

9. Some disclaimers, perhaps, are in order. First, I am a white, middle-class law profes-

against standardized testing that bear on law school admissions. I shall first say a few words about the quality of argument in each of the books and then proceed to the major issues of test bias and validity.

I. The Critics

It is tempting to dismiss Allan Nairn's report on ETS, "The Corporation that makes up minds," as *haute* yellow journalism.[10] Nevertheless, the book has been taken seriously in publications ranging from *Reader's Digest*[11] to *Fortune*[12] to *Science*,[13] and much of what it says is true enough. The products of corporations like ETS and American College Testing (ACT) do have tremendous impact on the lives and careers of many Americans.[14] So-called "intelligence" and "aptitude" tests[15] do stigmatize and tend to become self-

sor. This admission conveys a double-edged truth. It implies that I have been judged worthy under the regime of standardized testing and that I might be grateful to the testers for this. On the other hand, the fact that I have taken many standardized tests should help me empathize with "a consumer perspective and analysis of ETS as judge and gatekeeper for millions of young Americans." NADER REPORT at xi. Second, in some of my more impecunious moments, I have written a few LSAT questions. Third, as a former member and chairman of an admissions committee, I have used LSAT scores in making decisions about law school applicants. Finally, and perhaps most awkward for my claim of objectivity, I am a member of the Test Development and Research Committee of the Law School Admission Council (LSAC), a group that the *Nader Report* characterizes as "the client organization convened by ETS to sponsor the LSAT." *Id.* at 18. *But cf. id.* at 220 ("[T]his Council enjoyed a reputation as the most independent and probing of the ETS client organizations.") All of this, I suppose, goes to show that Nairn and his associates are right: the tentacles of ETS stretch everywhere. *See* p. 435 & note 10 *infra.*

10. The aspects of the *Nader Report* that prompt this characterization include a proclivity toward sensationalism in description and toward ascription of guilt by association. Consider, for example, the revelation that a hotel and conference center complex owned by ETS housed a meeting of "a secret society of European nobility, world leaders and multinational executives," which "featured a discussion of the neutron bomb." NADER REPORT at 39-40. Then there is the chapter linking ETS to Nazi policies of "eugenic" sterilization. *Id.* at 161-96. But the *ad hominem* flavor of the expose is perhaps best revealed by the glossary. This glossary is not a list of the technical terms or important concepts in psychological testing. Instead, the "glossary" is nothing less—or more—than a cast of characters, with such edifying entries as *"Carnegie Corporation of New York:* One of the founders of ETS." *Id.* at 548.

11. Brownstein & Nairn, *Tests That Can Cripple Carriers*, READER'S DIG., Mar. 1980, at 157.

12. Seligman, *Better Than Dice*, FORTUNE, Mar. 10, 1980, at 43.

13. Smith, *Nader Assails ETS*, 207 SCI. 508 (1980). Among other discussions of the study are Bourgeois, *Multiple-Choice Anxiety*, MACLEANS, Feb. 4, 1980, at 39; Fallows, *The Tests and the "Brightest": How Fair Are the College Boards?* ATLANTIC, Feb. 1980, at 37; Wynne, *The College Testing Controversy*, Wall St. J., Feb. 14, 1980, at 26, col. 4.

14. *See* NADER REPORT at 43-54.

15. Nader and his associates have a field day debunking inflated, and often dated, claims about the ability of tests to measure intelligence or mental capacity. Most educational psychologists concede that there are no "intelligence" or "aptitude" tests that mea-

fulfilling prophecies.[16] Misuse of test scores by persons who should know better is disconcertingly common.[17] Educational and employment decisions are often based on tests that have little or no demonstrated ability to predict long-term professional success[18] and that typically have a disproportionate impact on the poor, on blacks, on Hispanics, and on other groups.[19]

In itself, this might seem a sufficient indictment of the whole testing enterprise, and if all the *Nader Report* did was to reiterate these well-known facts, the only grounds for complaint might be packaging or price.[20] But the book brings so many other accusations against ETS that the organization begins to resemble the dia-

sure innate mental capacity. *See* E. HILGARD, INTRODUCTION TO PSYCHOLOGY 404 (3d ed, 1962) ("Intelligence is that which an intelligence test measures."); J. SATTLER, ASSESSMENT OF CHILDREN'S INTELLIGENCE 8 (rev. ed. 1974) (innate mental capacity can never be measured directly). All that existing tests can measure are limited sorts of achievements; and the less recently the skills in question have been acquired, the more likely it is that the test will be labeled an intelligence or aptitude test. *Cf.* Anastasi, *Abilities and the Measurement of Achievement*, in 5 NEW DIRECTIONS FOR TESTING AND MEASUREMENT 1 (W. Schrader ed. 1980) (cognitive tests measure "developed abilities").

Educators in law and medicine have adopted a simple expedient to circumvent the objection that the LSAT and the Medical College Admission Test (MCAT) do not measure professional aptitude. They let the letter "A" stand for "Admission" rather than "Aptitude." *But see* Law School Admission Council, 1980-81 Law School Admission Bulletin and LSAT Preparation Material 4 (describing test as measuring "mental abilities important in the study of law") [hereinafter cited as LSAT Bulletin]. This strategem possesses the important virtue of honesty, but it leaves the test users wide open to an obvious question: if the tests do not measure professional aptitude, why should they be used in determining admissions to professional schools?

16. *See* NADER REPORT at 5-27.

17. Until quite recently, the Department of Justice and the Federal Trade Commission used LSAT scores in hiring decisions. NADER REPORT at 232. That even those for whom the LSAT is designed, the law schools, sometimes place undue reliance on the scores, is shown by the fact that in 1978, the Law School Admission Council felt impelled to distribute a set of "cautionary policies," urging member schools not to set any kind of "passing" score on the LSAT as a prerequisite for admission to law school. Law School Admission Council, Cautionary Policies Concerning Use of I. The Law School Admission Test (LSAT) II. The Law School Data Assembly Service (LSDAS) III. The Law School Candidate Referral Service (LSCRS) 2 (1978) [hereinafter cited as Cautionary Policies]. The *Nader Report* provides more anecdotal evidence of misuse. The opening chapter consists mostly of "case histories"—tearjerkers, really—of students who were misjudged, by themselves or others, because of low board scores; and nine pages are devoted to the plight of an unnamed blind student who suffered retinal detachment and further impairment of his vision because unnamed law schools refused to waive their requirement (or were not asked to do so—the account is not clear on this point) that all applicants submit an LSAT score. NADER REPORT at 14-22.

18. *See* NADER REPORT at 76-81.

19. *Id.* at 110; *see* R. SAMUDA, PSYCHOLOGICAL TESTING OF AMERICAN MINORITIES 1-3 (1975).

20. The book costs thirty dollars, a hefty sum for a publication that lacks a useful index, is not typeset, and is bound in paper.

bolical corporate villain of the latest Trevanian thriller:[21] an insensitive, immoral, secretive multinational enterprise,[22] complete with its own dreadfully efficient security force and nearly omniscient computer network,[23] callously forcing a useless product upon unsuspecting consumers and ruthlessly crushing helpless opponents.[24] Although this may sound more like fiction than fact, it would be surprising if at least a few of the unflattering features of the portrait did not have some basis in reality. Charges of complacency, insensitivity, secretiveness and arrogance on the part of a major corporation that enjoys monopoly power have a certain plausibility, and ETS probably deserves a good, swift kick from time to time. The *Nader Report* does its utmost to deliver the blow, but its delivery is flawed by its unbalanced stance.

The National Conference of Black Lawyers (NCBL) report covers much of the same ground as the *Nader Report*, but it focuses specifically on the LSAT. Prepared by a longstanding critic of standardized testing, David White, under a grant from the National Institute of Education, this report also picks and chooses from the available research to support certain conclusions, but the argumentation is tighter, more sophisticated, and less infected with obvious errors and contradictions. The report includes an elabo-

21. TREVANIAN, SHIBUMI (1979).

22. Trevanian's implacably evil opponent is the "Mother Company," a consortium of major international petroleum, communications, and transportation corporations that controls the western world's energy and information, and dictates to the CIA and heads of state. *Id.* at 25. Likewise, Nairn's opponent, the Educational Testing Service, comprises "an international network with more outposts than the U.S. Department of Defense, and . . . extends . . . 'from Antarctica to Zaire.'" NADER REPORT at 29. Its leaders dine with assorted "dukes, counts and statesmen." *Id.* at 40. Its agents act "on instructions from the ETS central office" and move to "rendezvous points through the streets of Belgrade, Yugoslavia; Cape Town, South Africa; Natchez, Mississippi; Scranton, Pennsylvania; Moscow, Idaho; Kathmandu, Nepal; Morgantown, West Virginia; Stuttgart, Germany; Queens, New York; Benghazi, Libya; Malibu, California; Fort Wayne, Indiana; and Montivideo, Uruguay." *Id.* at 29.

23. Trevanian calls the computer system employed by his Mother Company the Fat Boy. TREVANIAN, *supra* note 21, at 15. Nairn tells us that ETS uses "The Wizard of OS." NADER REPORT at 28. Fat Boy is fed facts through "the constant work of an army of mechanics and technicians." TREVANIAN, *supra* note 21, at 15. The Wizard is tended by "[s]eventy technicians and a staff of IBM consultants [who] work three shifts daily." NADER REPORT at 28. Fat Boy contains "a medley of information from all the computers in the Western World . . . the most delicate information, and the most mundane. If you lived in the industrialized West, Fat Boy had you." TREVANIAN, *supra* note 21, at 15. The Wizard "coordinates the largest data bank of personal educational and psychological information in the world: it sorts files on more than thirty-two million persons from one hundred nations." NADER REPORT at 28.

24. Trevanian's Mother Company uses CIA assassins to do away with harmless opposition. TREVANIAN, *supra* note 21, at 299. ETS relies on the Washington, D.C., law firm of Wilmer & Pickering to cut into *pro se* litigants with "legal micro-surgery." NADER REPORT at 284.

rate intuitive analysis of several sample LSAT questions, which suggests that black-white cultural differences might affect the responses.[25] It also includes a criticism of law school teaching and grading practices.[26] It argues that LSAT scores, as opposed to college grades, have little value for admissions purposes,[27] and that undergraduate grades have much less of an adverse impact on blacks.[28] It urges that reliance on LSAT scores be reduced or eliminated, or that the scores of minorities be adjusted upward in some fashion.[29]

The Mexican American Legal Defense and Education Fund (MALDEF) report is more modest in scope and less radical in its call for reforms in law school admissions. Like the *Nader* and the NCBL reports, the *MALDEF Report* insists that the LSAT is racially and economically biased.[30] Nonetheless, accepting the view that " '[p]sychometricians can give us valuable tools, but they must not be misused,' "[31] it proposes several "alternative admissions models"[32] designed to make more places available to minority students while continuing to use LSAT scores. This portion of the report should prove particularly useful to those law school administrators, faculty, and students concerned with setting admissions policy.[33]

II. The Issues

Two fundamental criticisms of the LSAT emerge from the three books: first, the test is economically, culturally, and racially biased; second, the test is an inaccurate measure of the qualities law stu-

25. NCBL Report at 63-89.
26. *Id.* at 108-27.
27. *Id.* at 22-43, 100-01.
28. *Id.* at 101-03, 106.
29. *Id.* at 131-45.
30. The discussion of this point is confused, and at one point the report attempts to dismiss the whole issue of cultural bias as "obfuscatory." MALDEF REPORT at 19.
31. *Id.* at 39.
32. *Id.* at 39-61. The models are proffered as "alternatives" to a system that would admit only those students whose college grade point averages and LSAT scores, when combined according to a certain linear equation, would give them the highest predicted grades in their first year of law school. Most law school admissions officers would deny that admissions decisions are inexorably fixed by such a crude and mechanical process. *See* Evans, *Applications and Admissions to ABA Accredited Law Schools: An Analysis of National Data for the Class Entering the Fall of 1976*, in 3 REPORTS OF LSAC SPONSORED RESEARCH, 1975-1977, at 551, 622 (1977) (Report No. LSAC-77-1); *cf.* note 100 *infra* (describing admissions process that uses LSAT, grades, and other factors).
33. In exploring some of the alternatives to blind reliance on "the numbers," the *MALDEF Report*, in contrast to the other two reports, includes some original empirical research: a "pilot study" of applicants to three San Francisco Bay Area schools. MALDEF REPORT at 58-61. One feature of this research is criticized below. *See* note 100 *infra*.

dents and lawyers should have. If these criticisms were as strongly supported as the three reports suggest, testing should cease.[34] This section therefore analyzes these two criticisms. I conclude that the complaints against the LSAT are overstated, but that the reasons for relying on test scores to make selections among applicants who are all evidently capable of doing reasonably well in law school are less impressive than many admissions officers and faculty members assume.

A. *Test Bias*

Unfortunately, the clarity of the debate about test bias is diminished by the slipperiness of a short word like "bias," or for that matter, a long one like "discrimination."[35] Educational psychologists have spawned a variety of technical definitions of test bias, not all of which are consonant with the legal concepts of equal protection and nondiscrimination.[36] It is therefore helpful to analyze the issue by using familiar terms from employment-discrimination litigation before turning to the psychometric definitions of bias.

The criticism concerning test bias does not allege discrimination in the sense of disparate treatment. Schools that demand LSAT scores require them for almost all applicants.[37] Rather, the argument is that, like a law that prohibits the rich and the poor alike from sleeping under bridges or stealing bread, the test has a disparate impact on minority applicants. This effect is said to arise, first,

34. If, as the *Nader Report* claims, using the tests is on a par with consulting Ouija boards and rolling dice, "truth in testing" legislation is far too mild a palliative. In view of the adverse impact of standardized testing on blacks, the NAACP, among others, has called for a testing moratorium. R. SAMUDA, *supra* note 19, at 4-5. In evaluating the abolitionist position, it is well to distinguish between, on one hand, the use of "intelligence" tests, such as the WISC and Stanford-Binet, to place children reared outside the white, middle-class culture in special education classes for the "mentally retarded," and, on the other hand, the use of other standardized tests in college, graduate, and professional school admissions. This review focuses only on the latter issue. The *Nader Report*, by contrast, blurs the distinction between the two kinds of tests by viewing all standardized testing as a monolithic device to perpetuate class distinctions. *See* NADER REPORT at 117-19. For discussions of the distinctive cultural-bias and proper-use problems with children's "intelligence" tests, see, for example, Parents In Action On Special Education v. Hannon, 49 U.S.L.W. 2087 (N.D. Ill., July 7, 1980); J. SATTLER, *supra* note 15, at 25-49; Messe, Crano, Messe, & Rice, *Evaluation of the Predictive Validity of Tests of Mental Ability for Classroom Performance in Elementary Grades*, 71 J. EDUC. PSYCH. 233 (1979).

35. *See* Williams, *Preference, Prejudice and Difference—Racial Reasoning in Unfree Markets*, 3 REG., Mar./Apr. 1979, at 39.

36. *See* note 95 *infra*.

37. At most law schools, the requirement is waived for handicapped students. *See* Law School Admission Service, Operations Reference Book § 2, at 17 (1979) [hereinafter cited as LSAS Reference Book]; LSAT Bulletin, *supra* note 15, at 9. *But see* note 17 *supra* (discussing *Nader Report*'s account of plight of blind applicant).

because more affluent students have access to expensive cram courses, and second, because the phrasing and content of the LSAT questions make the correct answers more apparent to majority applicants of higher socio-economic status.[38]

The fact of disproportionate impact is undeniable. Oddly, the discrepancy in the mean scores of majority and minority applicants is rarely reported, but it appears that, on the average, whites score approximately 100 points higher than blacks, Hispanics, and Native Americans.[39] The explanation for this disparity, however, is less clearcut. Take the cram-course hypothesis advanced by the *Nader Report*.[40] It can by itself explain the bulk of the disparity only

38. Other reasons are occasionally offered. The NCBL Report, for example, suggests that "the speed requirement of the LSAT may have a discriminatory impact." NCBL Report at 47. The available evidence does not support, or conclusively disprove, this hypothesis. *See* Evans & Reilly, *The LSAT Speededness Study Revisited: Final Report*, in 2 REPORTS OF LSAC SPONSORED RESEARCH, 1970-74, at 191 (1976) (Report No. LSAC-72-3); Rindler, *Pitfalls in Assessing Test Speededness*, 16 J. EDUC. MEASUREMENT 216 (1979).

39. Although the *MALDEF Report* states that "ETS research conducted in 1973 established that there is a 133 point discrepancy between the average white and Black male LSAT scores and an 89 point discrepancy between the white and Chicano scores," MALDEF REPORT at 18, I have been unable to locate any ETS or LSAC study that provides such figures. My own analysis of the LSAT scores of 1129 applicants for admission to the College of Law at Arizona State University (ASU) this past year, however, is consistent with the *MALDEF Report*. Where the mean LSAT for the 1024 whites was 573, the mean figure for the 21 blacks, 62 hispanics, and 15 Native Americans was 457, 467, and 491, respectively. Of course, these figures pertain solely to one school and may differ somewhat from the national picture, but national studies released by LSAC also point to disparities in the vicinity of 100 points. Although the mean LSAT score is about 525, LSAS Reference Book, *supra* note 37, § 2, at 3, and 72 percent of all persons score in the 500 to 800 range, only 19 percent of blacks and 33 percent of Hispanics register scores in this range. Evans, *supra* note 32, at 604.

40. The *Nader Report* refers to commercial cram-courses to suggest that the poor are disadvantaged in the competition for high test scores and to attack the representations of the College Entrance Examination Board (CEEB) that short-term coaching can produce only negligible score gains on the Scholastic Aptitude Test (SAT). NADER REPORT at 96-109. Although commercial coaching establishments are estimated to do about $10,000,000 worth of business annually, FEDERAL TRADE COMM'N, STAFF MEMORANDUM OF THE BOSTON REGIONAL OFFICE OF THE FEDERAL TRADE COMMISSION—THE EFFECTS OF COACHING ON STANDARDIZED ADMISSION EXAMINATIONS 31 (Sept. 1978) [hereinafter cited as BOSTON STAFF MEMORANDUM], the inability of even the largest coaching schools with franchises across the nation to document advertising representations of dramatic score increases does not trouble Nairn or his staff, who see this practice as "another routine case of marginal enterpreneurs caught cutting corners to make a buck." NADER REPORT at 102.

The report is, however, on firmer ground in criticizing the College Board's claims. *See* Slack & Porter, *The Scholastic Aptitude Test: A Critical Appraisal*, 50 HARV. EDUC. REV. 154, 155 (1980). The materials distributed to persons who register for the LSAT contain the following representation: "There is no evidence available to LSAC, LSAS, or ETS that taking LSAT preparation courses improves an examinee's score on average or gives an advantage that cannot be attained by conscientious study of the LSAT preparation material in the *Bulletin*." LSAT Bulletin, *supra* note 15, at 17. The latter part of this claim may be correct, but the use of the phrase "no evidence" instead of the more accurate "no conclusive evidence" is misleading.

if three conditions hold: the fees for these courses, which range from $40 to $300,[41] exclude most minority applicants; most majority applicants enroll in the courses; and participation in the courses raises LSAT scores by an average of approximately 100 points.[42]

Substantial evidence that *any* of these conditions holds, however, is hard to find. The *Nader Report* relies primarily on a regional staff investigation by the Federal Trade Commission of twenty-one commercial coaching schools in four metropolitan areas.[43] This study did not address the first two points,[44] and its findings provide little support for the third condition. According to Nader's staff, the study found that graduates of one such course scored sixty points higher, on the average, than did persons who took the test without commerical coaching.[45] Apparently, comparable gains were not posted by other courses.[46] Moreover, the study did not adequately consider probable differences in preparation and motivation between the coached and uncoached students.[47] There is some indi-

41. NADER REPORT at 96.

42. These conditions are obviously interrelated. For instance, if coaching boosts scores by an average of 200 points, then fewer whites would need to enroll in cram-courses to produce the 100 point disparity.

43. BOSTON STAFF MEMORANDUM, *supra* note 40, at 31. Two SAT preparatory courses in metropolitan New York were also studied. *Id.* at 50.

44. A reexamination of the data collected by the Boston office on the two SAT courses reveals that minorities are not completely excluded from commercial SAT coaching schools. In fact, the proportion of minority students in the coached population is 10.3 percent—very close to the 11.3 percent figure for the uncoached group. BUREAU OF CONSUMER PROTECTION, FEDERAL TRADE COMMISSION, EFFECTS OF COACHING ON STANDARDIZED ADMISSION EXAMINATIONS, REVISED STATISTICAL ANALYSES OF DATA GATHERED BY BOSTON REGIONAL OFFICE OF THE FEDERAL TRADE COMMISSION 9 (Mar. 1979) [hereinafter cited as REVISED STATISTICAL ANALYSES]. The Bureau's definition of "minority," however, may include Asian Americans. Blacks do tend to cluster in the uncoached group, representing 6.9 percent of the uncoached, but only 3.7 percent of the coached students. *Id.* Furthermore, although student responses indicated that the parents of 49 percent of the uncoached group had incomes of less than $18,000, only 31 percent of the parents of students who obtained commercial coaching had similarly low incomes. *Id.* at 8.

45. NADER REPORT at 102. The published version of the *Boston Staff Memorandum* does not report such a result, but the *Nader Report* indicates that this portion of the staff's study was "edited out [by the Commission] and . . . [has] never before been released to the public." *Id.* at 106. In fact, the published version of the study concludes that, whereas coaching is "dramatically effective" for the SAT, the effect on LSAT scores is sometimes "marginal." BOSTON STAFF MEMORANDUM, *supra* note 40, at 157. The regional staff, however, argues that the lack of susceptibility of the LSAT to commercial coaching is merely "apparent" and that such susceptibility might be demonstrated by further research. *Id.* at 157.

46. Not having seen the unpublished portion of the *Boston Staff Memorandum*, I can only assume that the failure of the *Nader Report* to mention gains for any other LSAT cram-courses is deliberate. By my count when reading the Boston study, between 14 and 20 LSAT courses were investigated. *See* BOSTON STAFF MEMORANDUM, *supra* note 40, at ix-x, 37-40, 50. Yet the Nader staff states that the sixty-point gain was posted in one "of the two LSAT courses studied." NADER REPORT at 102.

47. If the persons who enroll in commercial coaching schools were better prepared for

cation that, in view of such differences between the two groups, little of the sixty-point discrepancy can be attributed to participation in the course itself.[48] Given such negative findings, it seems doubtful that the cram-course hypothesis can explain more than a very small portion of the disparate impact of the LSAT.[49]

Nevertheless, concern over minority access to commercial cram-courses is only part, and perhaps a small part,[50] of the complaint that the LSAT is a disciminatory instrument. It is also possible that the questions are biased against minorities. The NCBL Report develops this thesis through a lengthy examination of selected ques-

the LSAT to begin with or were more highly motivated, the sixty-point difference between the mean LSAT score for coached and uncoached candidates might not reflect the effect of coaching at all. *See* Bureau of Consumer Protection, Notice to Recipients of the Boston Regional Office Report on the Effects of Coaching on Standardized Admission Examinations (May 1979). The FTC released the Boston study to the public with the admonition that the study " 'has several major flaws in the data analysis, making the results unreliable,' " NADER REPORT at 106 (quoting from page stapled to cover of BOSTON STAFF MEMORANDUM, *supra* note 40), and with the further statement that "the Commission specifically believes that some of the conclusions in the study are not supported by the evidence obtained in the investigation," BOSTON STAFF MEMORANDUM, *supra* note 40 (title page).

The *Nader Report* dismisses the Commission's reservations as purely "technical." NADER REPORT at 104. Although we are told that the Boston Office "accumulated an information base unprecedented in the history of coaching studies," *id.* at 101, Nader concludes that the inability of the Boston office to demonstrate that coaching *caused* any gains on the SAT or the LSAT arose only because "there were not enough data available," *id.* at 104. Indeed, it is ETS that emerges as the real villain. It seems that ETS beseiged the Commission with methodological criticisms of the Boston report, and the reticence of the FTC to endorse its staff's report represented, in part, "covering the Commission's flanks from ETS disapproval." *Id.*

48. When the FTC's Bureau of Consumer Protection attempted to correct for at least some of the differences in the composition of the coached and uncoached groups that took the SAT, it found that the gap in mean SAT scores between the two groups disappeared or diminished greatly. *See* REVISED STATISTICAL ANALYSES, *supra* note 44, at 20, 29.

49. This conclusion by no means implies that coaching is never effective, that it might not be of some benefit to minorities, or that egalitarian concern about access to cram-courses is entirely misplaced. For example, analysis of the effectiveness of commercial cram-courses is complicated by the fact that a number of undergraduate colleges offer preparation for the LSAT. BOSTON STAFF MEMORANDUM, *supra* note 40, at 41. Furthermore, it may be that certain kinds of persons benefit from extended drill or advice on how to take multiple choice tests whereas others do not. SAT data, for example, suggest that "overachievers"—those who do worse on standardized examinations than would be predicted from their grade point averages or class ranks—find cram-courses beneficial while most other students do not. *See* REVISED STATISTICAL ANALYSES, *supra* note 44, at i-ii. For a comprehensive review of the SAT coaching studies, see Slack & Porter, *supra* note 40.

Of course, to the extent that low-income or minority students have access to inexpensive or free preparatory courses, or can achieve similar results by a conscientious program of self-study based on preparatory materials supplied by testing organizations, the egalitarian concern is mitigated. In this connection, the *Nader Report*'s call for an increase in official advice to students on how to be "test-wise" is quite appropriate. *See* NADER REPORT at 92-95.

50. *See* note 40 *supra*.

tions from the LSAT Bulletin distributed to candidates. The NCBL charges that the phrasing or content of a number of the Bulletin's sample questions displays insensitivity to minority-group traditions, ignorance of minority-community values, assumptions contrary to those of minority-group members, prejudicial stereotypes about minority-group members, disregard of bilingual concerns, unnecessary confusion with black standard English, and ignorance of the history of black culture.[51] These are serious accusations. Before accepting them at face value, it would be useful to know whether the sample questions flagged by the NCBL Report do in fact prove more difficult for minority candidates. This is a question that cannot be resolved by armchair research. It could have been investigated, however, by administering the sample test to matched groups of black and white students.[52]

In fact, a number of statistical studies have examined the possibility that questions actually used on the LSAT are biased in this sense.[53] The most recent involved 195 LSAT questions and thousands of minority students.[54] Thirty-eight questions appeared to be harder for either blacks or Hispanics,[55] but nearly twice this number proved *easier* for these groups.[56] Thus, the NCBL Report may be correct in suggesting that certain questions are more difficult for blacks than for equally able whites and that some of these questions

51. *See* NCBL Report at 63-88.

52. Other research designs are also possible. *See* Schmeiser & Ferguson, *Performance of Black and White Students on Test Materials Containing Content Based on Black and White Cultures*, 15 J. EDUC. MEASUREMENT 193 (1978).

53. *E.g.*, Swineford, *Comparisons of Black Candidates and Chicano Candidates with White Candidates*, in 2 REPORTS OF LSAC SPONSORED RESEARCH, 1970-74, at 261 (1976) (Report No. LSAC-72-6).

54. L. Wightman, Study of LSAT Item Performance for Different Subgroups (Oct. 1979 draft). The questions were all taken from one form of the LSAT then in use. Students, taking the test in December 1976, were divided into five categories: black men, black women, Chicano men and women, white men, and white women. *Id.* at 6. The number in each category ranged from 915 for the Chicano group to 1,874 for the black men. *Id.*

55. The difficulty of each question was determined from the item-characteristic curves of each question for each ethnic group. An item-characteristic curve attempts to display the probability of a correct response as a function of the ability of the persons taking the test. *See* Hambelton & Cook, *Latent Trait Models and Their Use in the Analysis of Educational Test Data*, 14 J. EDUC. MEASUREMENT 75, 78-81 (1977). For discussions of the relative merits of the various statistical techniques for detecting bias in individual questions, see Ironson & Subkoviak, *A Comparison of Several Methods of Assessing Item Bias*, 16 J. EDUC. MEASUREMENT 209 (1979); Rudner, Getson, & Knight, *A Monte Carlo Comparison of Seven Biased Item Detection Techniques*, 17 J. EDUC. MEASUREMENT 1 (1980).

56. Seventy-one questions were found to be easier for one minority group or another. For each of the minority groups, the number of easier questions exceeded the number of harder ones. L. Wightman, *supra* note 54 (table 15). For white women, however, there was an excess of three harder questions over easier ones. *Id.*

should be eliminated;[57] but it would not seem that the test as a whole is biased in this sense.[58]

This leaves us in an awkward spot. If neither the existence of commercial cram-courses nor the cultural bias seen in the formulation of specific questions provides a satisfactory account of the disparity in majority-minority LSAT scores, what does? The suggestion that the disparity has a substantial genetic component is not likely to be greeted with enthusiasm.[59] As a purely scientific matter, the evidence on racial-ethnic differences in performance on intellectual tests can accommodate a wide range of views of the genetic hypothesis.[60] Fortunately, this perennial and divisive debate need not be resolved here. As the *MALDEF Report* points out, the groups that do poorly on the LSAT are the very same groups that have been victimized by de jure or de facto discrimination in education.[61] A footrace may go to the swiftest, but those who have not benefited from a meaningful training program are not likely to finish in the winner's circle, irrespective of any genetic component of being fleet of foot. On this view, the score disparities mirror differ-

57. It does not necessarily follow that because this study found a question to differ in difficulty between two groups, that question should be eliminated. Due to the large sample sizes, minute differences in difficulty can be statistically significant. For illustration, suppose that difficulty is measured on a scale ranging from zero to one and that the difficulty of a question for blacks is found to be 0.5001, compared to 0.5000 for whites. If these numbers had been derived from only a handful of observations on black and white test-takers, one might think that the difference of 0.0001 would not show up if the performances of a much larger number of students had been examined. Because it is risky to generalize from such small samples, one might say that the difference is not *statistically* significant. On the other hand, if the difference had been derived from many thousands of observations, it would be harder to dismiss as a sampling error. Instead, one might conclude that the difference is statistically significant; it probably reflects a true difference between the general populations of black and white test-takers. Nevertheless, this difference of 0.0001 might be of little practical consequence. It might have so slight an impact on actual scores as to be negligible, and it might be balanced by another question that is slightly easier for blacks. Thoughtful statisticians express this idea by distinguishing between "statistical" and "practical" significance. *See* D. MOORE, STATISTICS: CONCEPTS AND CONTROVERSIES 292 (1979).

58. It is not clear from the study, however, whether the test is on balance easier for minorities when the biased questions are weighted by the degree of bias.

59. The *MALDEF Report* calls this suggestion "eugenicist-type." MALDEF REPORT at 19.

60. *See, e.g.*, J. LOEHLIN, G. LINDZEY, & J. SPUHLER, RACE DIFFERENCES IN INTELLIGENCE 238-39 (1975) ("A rather wide range of positions concerning the relative weight . . . [of biases and inadequacies of tests, of environmental factors, and of genetic factors] can reasonably be taken on the basis of current evidence.") Bodmer & Cavalli-Sforza, *Intelligence and Race*, SCIENTIFIC AM., Oct. 1970, at 19-29 ("currently available data are inadequate to resolve . . . [the] question in either direction" of whether "there could be a genetic component in the mean difference in IQ between races").

61. MALDEF REPORT at 19-20; *see* Brief Amicus Curiae of the Law School Admission Council at 15, Regents of the University of California v. Bakke, 438 U.S. 265 (1978) (one examination shows that more than 60% of potential minority law-school candidates attended unlawfully segregated public schools).

ences in educational achievement, which reflect fundamental social inequities rather than variations in innate abilities.

To be sure, this analysis—which is the ETS party line[62]—presupposes that tests like the LSAT measure certain educational attainments with at least moderate accuracy.[63] In addition, it does not completely refute the charge of test bias, for it is also possible that the LSAT systematically understates the achievements of low-scoring groups or overstates those of high-scoring groups. If this were so, differences in group means would be observed, but these differences would exaggerate the disparities in achievement that are actually present. There is a statistical method for examining this possibility, and it has been applied to the LSAT. To describe this procedure, it is best first to explain the techniques for assessing the test's general accuracy. It is to this issue, which educational psychologists usually call test validity, that we now turn.

B. *Test Validity*

Standardized tests are measuring instruments. Like college grades, they measure something about intellectual performance. To understand how we can tell whether the measurements they provide are accurate, an extended analogy may be useful. Consider a mechanical device, say a bathroom scale, that measures some physical property like weight. The specific scale I have in mind keeps insisting that I am overweight. To check its accuracy, I might try putting the same heavy book on it every morning. If the measured weight varied dramatically from day to day, I might decide that the scale was not working correctly. In the literature on testing, the comparable measure is called reliability, and the reliability of the LSAT is said to be quite high.[64] The fact that my bathroom scale gives consistent measurements, however, does not mean

62. *See* Turnbull, *Foreword* to R. SAMUDA, *supra* note 19, at vii, ix (Foreword by President of ETS).

63. This supposition is scrutinized below. *See* pp. 443-53 *infra*.

64. The LSAS Operations Reference Book, *supra* note 35, § 2, at 7, states:

> An "internal consistency" reliability coefficient is computed for each edition of the LSAT and is an estimate of the degree to which individuals would keep the same relative standing in a group, if equivalent forms of the test were administered to all members of a representative group with effects of practice and fatigue removed . . . .
>
> The average reliability of the scores on recent forms of the LSAT is .92 for a group of candidates very similar to the total group of candidates tested. This is certainly high enough to justify the use of the scores in individual counseling and prediction.

For a clearer discussion of the methods of calculating reliability, see A. ANASTASI, PSYCHOLOGICAL TESTING 78-94 (3d ed. 1968). For a more involved treatment, see Cudeck, *A Comparative Study of Indices for Internal Consistency*, 17 J. EDUC. MEASUREMENT 117 (1980).

that it provides correct measurements. So, too, the finding that the LSAT is "reliable" does not prove that it is accurate.

Recognizing that internal consistency is not sufficient, I might try to compare the readings given by my scale with those of other scales. This approach, though, is less useful with the LSAT. Supposedly, there are no other tests that measure exactly what the LSAT measures.[65] Moreover, even if all the bathroom scales I examined were in agreeement, it could still be that they were all wrong. There might have been some uniform defect in their manufacture or some recent environmental change that affected the whole lot.

I might next decide to take my scale apart to look for a defect. If my knowledge of physics and my mechanical acumen were adequate, I might be able to assure myself that all is well. In testing jargon, I would be looking for facial validity. The NCBL Report does a fair amount of this. For example, questions involving quantitative comparisons are criticized for lacking facial validity,[66] and even those involving "principles and cases" are challenged on the ground that they "do not simulate the actual legal practice of attorneys."[67] Although these particular criticisms seem somewhat over-

65. It would make little sense to use college grades (another measure of potential for successful law study or practice) as a measure of the accuracy of the LSAT. The LSAT is intended to measure something that is not measured or is measured less adequately by college grades. Nevertheless, the assertion that other standardized examinations are unsuited to law school admissions is probably false. *See* Schrader & Pitcher, *The Advanced Tests of the Graudate Record Examinations as a Predictor of Law School Grades*, in 2 REPORTS OF LSAC SPONSORED RESEARCH, 1970-74, at 47 (1976) (Report No. LSAC-70-5); *cf.* Slack & Porter, *supra* note 40, at 167-69 (achievement tests predict college academic performance better than SAT).

66. NCBL Report at 56. The logic seems to be that because the quantitative comparisons sometimes involve mathematical operations such as squaring numbers or taking square roots, and because few lawyers perform such mathematical manipulations in advising clients, writing briefs, or their other tasks, the quantitative questions lack content validity. This seems too narrow a view of facial validity. The ability to perform mathematical operations may not be tapped directly in the study or practice of law, but the ability to work with mathematical and logical operators may be evidence of the ability to think clearly. As Professor Lempert reminds us, "[w]hatever enables lawyers to think more clearly is of practical importance." Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1056 (1977). In other words, the LSAT may have "construct validity" even where it lacks "content validity." *See* Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.5(B) (1979).

67. The reason for the NCBL Report's criticism is that "a test which considers only one response correct does not fit the rationale of a legal system which is grounded on an adversary model in which both sides are represented by competent counsel." NCBL Report at 58.

Although questions that are phrased in terms of "principles and cases," "issues and facts," or the like, might obviously be thought relevant to legal study and practice, the use of such questions poses a dilemma. If the questions use standard legal terms with their conventional meanings, then these questions can be attacked as unfairly advantaging per-

drawn,[68] even more compelling demonstrations of the absence of facial validity should not be decisive. No argument about the relevance of the questions in the test to the measurement of certain qualities could be fully persuasive without some good theory of what legal "aptitude" is and how tests work. No such theory is currently available.

If my physical inspection of the bathroom scale was not entirely convincing, what more could I do to test the scale's accuracy? I might weigh some convenient objects and then use those weights to make predictions as to how the objects will behave when put into a cylindrical bucket and floated in my bathtub. If I knew the density of the water, the area of the bucket's base, and the weight of the bucket, I could use Archimedes' principle to calculate how deep the bucket with each object would be submerged. If, on performing the experiment, my observations of the submerged depths matched my predictions, then I might conclude that the scale works pretty well after all. In the *lingua franca* of psychometrics, this experiment would be an investigation of predictive validity.

In employment-discrimination cases, predictive validity translates into job-relatedness. Even if a test has a disparate impact on a protected class, scores can be used to exclude candidates for hiring or promotion as long as the test predicts job performance with reasonable accuracy.[69] Although law schools are not employers of

sons who have been exposed to legal materials. *See id.* at 59. On the other hand, if the terms are defined to have special meanings for the purposes of the test, the questions can be criticized as unfairly confusing individuals with prior legal knowledge. In the end, it may be that the old favorites, such as "reading comprehension," are the best. Surely it would be a narrow view of facial validity that would deny the relevance of questions that test whether an individual can read a few paragraphs and summarize the main ideas.

68. Why should it be necessary to "simulate . . . actual legal practice" in order to see whether aspirants to the bar can acquire the skills they will need to achieve professional success? If only such a simulation could provide facial validity, then almost everything in a student's history (including his grades, almost all work experience, almost all letters of recommendation, and almost all other academic and social achievements) would lack facial validity.

69. There is some confusion over the applicable evidentiary standards in employment-discrimination cases. *See* Friedman, *The Burger Court and the Prima Facie Case in Employment Discrimination Litigation: A Critique*, 65 CORNELL L. REV. 1 (1979). The leading cases are Griggs v. Duke Power Co., 401 U.S. 424 (1971), and Washington v. Davis, 426 U.S. 229 (1976). *Griggs v. Duke Power Co.* is usually read to require a showing of predictive validity in an action brought under Title VII whenever disparate impact is apparent. *See* Comment, *Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under Title VII*, 91 HARV. L. REV. 793, 793 (1978). *Washington v. Davis*, in contrast, is generally understood not to require such validation in equal-protection cases unless there is also distinct evidence of discriminatory intent. *See* Perry, *The Disproportionate Impact Theory of Racial Discrimination*, 125 U. PA. L. REV. 540, 543-44 (1977). Although this distinction between the evidentiary standards in Title VII litigation and those in equal-protection cases finds support in the

their students, it is instructive to consider how they might defend their use of the LSAT if Title VII standards were applicable.[70] Among other things, they might want to contend that the LSAT has predictive validity. But exactly what is the LSAT supposed to predict? If success as a lawyer is the appropriate criterion, then no direct demonstration of predictive validity is available. To be sure, some current research is seeking quantified measures of professional success, and it may even generate some information about the value of the LSAT as a prognosticator of career achievements.[71] But I, for one, would be quite surprised if the LSAT proves to be more than marginally correlated with success in the practice of law. After all, the test is administered early in a lawyer's professional life, and even its most dogmatic defenders would have to concede that it measures only one or two of the qualities necessary for outstanding professional accomplishments.

Nevertheless, to conclude on the strength of this argument that the LSAT is devoid of predictive validity would be too hasty. The same analysis would condemn college grades and many other factors commonly considered by admissions personnel. For instance, undergraduate grading practices have an adverse impact on blacks and Hispanics.[72] Because, as the *Nader Report* puts it, "the validity of grades in general as predictors of later success has not been upheld by evidence,"[73] it would seem that our Title VII analysis

Court's opinion in *Washington v. Davis*, it is not clear that the distinction is appropriate. The cases can be interpreted to create a uniform standard under which a prima facie case of discrimination is established by a showing of *both* disparate impact and questionable facial validity. Under this approach, the LSAT might have sufficient facial validity to avoid the need for formal validation. For present purposes, however, it is more revealing to ask whether the sort of predictive validity that is sometimes demanded in Title VII cases can be shown.

70. The *Nader Report* seizes upon an idea tentatively mentioned by Professor Millard Ruud, who once asked whether a law school might be liable under Title VII, as if it were an employer, because admissions decisions affect employment opportunities. NADER REPORT at 230-32 (reporting remarks made by Ruud at 1971 conference). This theory sparkles with originality, but no attorney familiar with Title VII litigation would be surprised to learn that "[e]ight years after *Griggs*, no such case is known to have been brought." *Id.* at 232.

71. *See* Carlson & Werts, *Relationships Among Law School Predictors, Law School Performance, and Bar Examination Results*, in 3 REPORTS OF LSAC SPONSORED RESEARCH, 1975-1977, at 211, 212-13 (1977) (Report No. LSAC-73-2).

72. National data indicate that 40 percent of white candidates for admission to law schools in 1976 had undergraduate grade point averages of 3.25 or better. In contrast, only 13 percent of black applicants and 22 percent of Hispanic students posted averages in this range. Evans, *supra* note 39, at 604. All the books reviewed here minimize the significance of these disparities. *But see* note 100 *infra* (discussing possible significance of disparities).

73. NADER REPORT at 80. The *Nader Report* cites a 1965 American College Testing research report that reviewed 46 studies, which "defined and measured professional achieve-

precludes reliance on college grades as well as LSAT scores.[74] To climb out of the hole we seem to have dug ourselves into, we can shift the focus from on-the-job performance, which is difficult to measure,[75] to success in a training program, which is more tractable.[76] If we can assume that a law degree and high grades in law school are sensible things for people who hire lawyers to look at,[77] then we can use success in the training program—in the law school—as a proxy for professional success. If, in addition, we use the grade point average obtained at the conclusion of the first year of law school (FYA) as a measure of success in law school,[78] then some relatively simple quantitative analysis of predictive validity becomes feasible.

With first-year grades as the criterion, how accurate is the LSAT? The *Nader Report* loudly proclaims that the test is essentially worthless: no better than "random prediction with a pair of dice" in all but thirteen percent of the cases.[79] To examine this dramatic claim, I compared two methods for predicting the FYA for students in the most recent first-year class at Arizona State University. First, I made predictions by consulting a table of random numbers. Second, I used an elementary statistical technique to derive (from the various LSAT scores and course performances of students in

ment by a multitude of standards, covered eight categories [business, teaching, engineering, medicine, scientific research, and miscellaneous others, and concluded] . . . 'that college grades bear little or no relationship to any measures of adult accomplishment.' " *Id.* (quoting D. Hoyt, author of ACT report) (footnote omitted).

74. *But see* note 69 *supra* (suggesting that use of LSAT may be legitimate under Title VII even without proof of predictive validity).

75. *See* Lerner, *Washington v. Davis: Quantity, Quality and Equality in Employment Testing*, 1976 SUP. CT. REV. 263, 291 ("criterion validity . . . becomes awesomely complex or fraught with danger when applied to complex, varied jobs").

76. In Washington v. Davis, 426 U.S. 229 (1976), the Supreme Court held that Title VII's requirement that employment tests be job-related could be satisfied by proof that the test validly predicted performance in a training course. *Id.* at 248-52. This result is controversial. For a vehement defense of the outcome of the case, see Lerner, *Employment Discrimination: Adverse Impact, Validity, and Equality*, 1979 SUP. CT. REV. 17.

77. This proposition is difficult to prove, or disprove, rigorously. *See* Lerner, *supra* note 76, at 18-19.

78. Even if it is agreed that academic performance in law school is an appropriate criterion for measuring the LSAT's predictive validity, the use of first-year grade point averages can be challenged. There are three principal arguments for using first-year grades. First, they permit validation on the basis of data that are only one year old instead of two or three. Second, they are probably highly correlated to cumulative averages at graduation. Third, most students who depart from law school for academic reasons do so in their first year, so that prediction of first-year performance seems especially important.

These arguments should in no way preclude the investigation of other criteria for evaluating predictive validity. For instance, clinical coursework is an important component of law-school performance, and it might be enlightening to know how the LSAT's value in predicting performance in these courses compares to its validity in predicting FYA.

79. NADER REPORT at 65.

the previous year's entering class) a simple equation for predicting FYA on the basis of the LSAT alone. I discovered that the true picture was almost exactly the opposite of what the *Nader Report* claims. For 113 out of 136 students—in eighty-three percent of the cases—predicting performance from the LSAT was superior to random prediction.

This result might suggest that the *Nader Report*'s conclusions are tendentious if not dishonest. The report's derivation of the thirteen percent figure, however, indicates that its authors are merely confused about the statistical method routinely used to assess predictive validity. This procedure, which goes by such forbidding names as "linear regression" and "least squares" fitting, is really quite simple. To expose the flaws in the report's account, a quick explanation of the method is in order. Imagine that we are asked to predict how each student accepted at law school *X* will do in his or her first year. If we have no information on the grades of former first-year students, we may as well guess at random. If we learn, however, that the mean FYA last year was, say 3.18, we probably can do better by merely predicting that every student will have a 3.18 this year.[80] To see just how much better predicting by the mean is than random guessing, we could use the following procedure: for each student, check the FYA at the conclusion of the year, subtract this number from our guess for that student, and square this difference; then add up these squared deviations from our guesses.[81] If we follow this method first using our random guesses and then using our 3.18 guess, we will arrive at two numbers, each indicating the overall accuracy of the respective methods of prediction. The smaller the number, the better the prediction.

If we have more information on past performance, we should be able to do even better in this "least squares" sense. Suppose, for instance, that last year's class consisted of ten persons whose LSAT scores and FYAs were as follows:[82]

LSAT:	450	493	555	567	610	618	635	660	691	729
FYA:	3.00	1.93	2.83	3.71	3.24	3.05	2.91	3.53	3.85	3.75

To visualize what these numbers reveal about the relationship between LSAT scores and FYA, we can draw a picture (sometimes

80. *See* Efron & Morris, *Stein's Paradox in Statistics*, SCIENTIFIC AM., May, 1977, at 119.

81. The differences are squared to prevent an instance of overprediction, which would have a positive sign, from cancelling an instance of equally mistaken underprediction, which would have a negative sign. Other techniques could also be used to circumvent this problem. *See* note 83 *infra*.

82. The mean for these FYAs remains 3.18.

called a scattergram) in which FYA is plotted against LSAT scores. With this graph, we can search for the line that best fits the data. For example, our previous approach of predicting by the mean is tantamount to drawing a straight line through the cloud of data points at the height of the mean. As the diagram below suggests, we can do better by drawing a line that slopes at a different angle and that intercepts the vertical axis at a different point. In particular, the line sketched below (commonly referred to as the regression line) can be shown to minimize, relative to all other possible straight lines, the sum of the squares of the vertical distance between each data point and the line itself.[83]

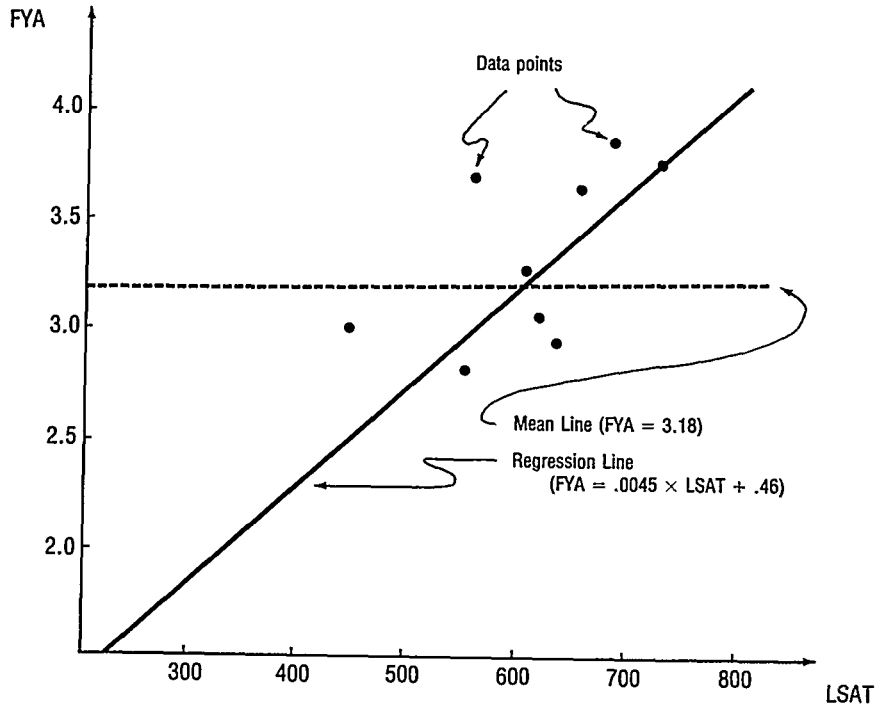

FIGURE 1: REGRESSION LINE AND MEAN LINE FOR HYPOTHETICAL CLASS AT LAW SCHOOL X

83. The proof involves elementary calculus. *See* A. EDWARDS, AN INTRODUCTION TO LINEAR REGRESSION AND CORRELATION 24-26 (1976). This technique of least-squares regression can be traced back to Newton.

The criticisms of the LSAT that are reviewed here do not question the assumption that the line that has the "best" fit is the one that minimizes the sum of the *squared* differences. Yet it is not clear that the costs of mistakes in predicting grades increase as rapidly as the square of the error in the predictions. Consequently, it might be argued that some other criterion would be more appropriate. For example, one might seek the line that minimizes

The method we used to compare the random guess with the 3.18 guess can now be employed to measure how well the regression line fits the data. We simply calculate the sum of the squared vertical distances between the data points and the line, and compare this number to the analogous sum for the horizontal line that depicts the mean. In this example, the former quantity is 1.62, and the latter is 2.99. Thus, the regression line has reduced the squared errors of prediction by an amount of 1.37 (2.99 − 1.62 = 1.37). In other words, the sloping line "explains" the proportion 1.37/2.99, or 46 percent, of the total variation from the mean. This proportion (conventionally denoted r^2) reports to what extent the regression-line method represents an improvement over prediction according to the mean; and, of course, this least squares prediction represents an even greater advance over random methods.

With this understanding of the most commonly used method for quantifying predictive validity, let us return to the thirteen percent figure cited in the *Nader Report*. This number, apparently, is derived from a report of the Carnegie Council on Policy Studies in Higher Education. That report does not present a figure for r^2. Instead it states that the median "validity coefficient" for the LSAT was .36.[84] This coefficient measures the correlation between FYA and LSAT scores.[85] Squaring it, the *Nader Report* obtains the figure of thirteen percent. It is true that the square of the correlation coefficient is numerically equal to r^2 as defined above. However, in stating that this number represents the percentage of cases in which the LSAT is no better than rolling dice in predicting FYA, the report makes no less than three elementary mistakes. First, it treats a median figure as if it pertained to each law school. By definition, half the schools will achieve better results than this median, and their use of the LSAT is not impeached by this generalized at-

the sum of the absolute values of the observed deviations. Even so, I would guess that techniques based on such loss functions would not yield a drastically different picture of the accuracy and fairness of the LSAT.

84. *See* NADER REPORT at 417 n.15, 422 n.32.

85. More precisely, the number is known as a zero-order Pearson product-moment correlation coefficient. J. GUILFORD & B. FRUCHTER, FUNDAMENTAL STATISTICS IN PSYCHOLOGY AND EDUCATION 83 (5th ed. 1973). If LSAT scores and FYAs were perfectly correlated, so that an increase of Δx points in the LSAT always corresponded to an increase of Δy points in FYA, the value of this coefficient would be one. If differences in LSAT scores bore no relation to differences in FYA, its value would be zero.

Nairn and his associates mistakenly state that the correlation coefficient involved in the ETS validity studies is a Spearman, or rank-order, correlation coefficient. Even if the .36 figure were a rank-order correlation, the interpretation of its square offered in the *Nader Report* would still be wrong. *See id.* at 284.

tack.[86] Second, this derivation assumes that r^2 can be directly compared with random prediction. It cannot. For large samples, the number gives the proportion of the variance explained by using the regression line instead of the mean to make predictions.[87] Because the mean is itself an improvement over rolling dice, r^2 understates the superiority of prediction based on the LSAT over prediction using random methods. Third, even as an indication of improvement over the mean, the statistic has no simple connection to the number of instances in which one method works better than another. Because r^2 is calculated by adding up the squared errors of prediction, a few large errors may yield the same number as would a great many smaller errors. That is, the magnitude of r^2 is determined not by *whether* the individual predictions according to one method come closer to the mark, but rather by *how much* closer these predictions come.[88]

In sum, the LSAT cannot be so easily dismissed as devoid of predictive validity. In fact, at most law schools it seems to be at least as accurate a predictor as undergraduate grades.[89] Considering Na-

86. Anecdoctal information indicates that the LSAT consistently predicts better for some schools than for others. The reasons for differences in validity from school to school have not been explored systematically, and the possibilities are numerous. For example, some schools may have anonymous grading in the first year, some may use a wider range of grades, some may have an entering class with widely dispersed LSAT scores, and some may report FYAs of students who flunk out in the first semester. All these factors should tend to produce higher correlation coefficients, because they would tend to produce more discriminating measurements of FYA or to curtail the "range-restriction" effect described below. *See* note 88 *infra*.

87. *See* pp. 449-50 *supra*; A. EDWARDS, *supra* note 83, at 43-44.

88. The statistic r^2 is a measure of overall fit, but there is no single number that adequately measures the accuracy of individual predictions. Probable accuracy decreases as one makes predictions for students whose LSAT scores are increasingly distant from the mean. *See* Binder, *Considerations of the Place of Assumptions in Correlational Analysis*, 14 AM. PSYCH. 504, 509 (1959).

89. The *Nader Report*'s source states that the median correlation of undergraduate grade point average with first year law school average is only .25. Willingham & Breland, *The Status of Selective Admissions*, in SELECTIVE ADMISSIONS IN HIGHER EDUCATION 65, 237 (Carnegie Council 1977). Yet the *Nader Report* does not announce prominently that predicting law school performance by college grades is a "respectable fraud," no better than rolling dice in 6 percent (approximately .25 squared) of the cases. Instead, the report observes in a footnote that "the relative predictive ability of grades and scores swings back and forth from one time period to another," NADER REPORT at 66 n.*, and it acknowledges the 6 percent figure only in another footnote two chapters later. *Id.* at 459 n.9. For evidence in support of this oscillation theory, the report defers to "attorney David White." *Id.* The NCBL Report, prepared by White, theorizes that as law schools place more and more emphasis on the LSAT, they will select students whose LSAT scores are virtually identical. This restriction of range, White reasons, will cause the measured validity of the LSAT to diminish, because predictions will then be based on negligible differences in LSAT scores. As this happens, law schools will return to attaching more importance to undergraduate grades, leading to a restriction of the range of this predictor, and to the subsequent

der's earlier achievements and his obvious talents, it is painful to read this chapter of a report published under his name. Although the accuracy of the LSAT—even at the limited task of predicting first year grades—is nothing to write home about,[90] it is far superior to dice.

reemergence of the LSAT as the apparently more valid predictor. NCBL Report at 92-95.

The phenomenon of range restriction is well known to statisticians. *See, e.g.*, Roe, *The Correction for Restriction of Range and the Difference Between Intended and Actual Selection*, 39 EDUC. & PSYCH. MEASUREMENT 551 (1979). But if the oscillation theory is to apply to the median correlation measured in 1976, two conditions must hold: the preponderance of the law schools must oscillate in phase, and undergraduate grades must show a more restricted range than the LSAT in the 1976 studies. The NCBL Report points to no data indicating that these conditions hold. To the contrary, a rough inspection of the evolution of correlations involving LSAT scores and undergraduate grades shows that the median correlation of LSAT scores with FYA has consistently exceeded that of undergraduate grades. Schrader, *Summary of Law School Validity Studies, 1948-1975*, in 3 REPORTS OF LSAC SPONSORED RESEARCH, 1975-1977, at 519, 530 (1977) (Report No. LSAC-76-8).

No adequate explanation for this result appears in the *Nader* or NCBL reports. One possibility is that "grade inflation" has restricted the range of grades in the applicant population, undermining the predictive validity of grades. *See* Singleton & Smith, *Does Grade Inflation Decrease the Reliability of Grades?* 15 J. EDUC. MEASUREMENT 37 (1978). This explanation, however, would not account for the success of the LSAT in pre-inflationary times. A more likely explanation is that the correlation of grades across the full spectrum of undergraduate institutions and of departments within institutions ignores important distinctions, reducing the measured validity. As the *Nader* and NCBL reports note, most of the previously published efforts to adjust grades according to the college attended have not been very successful. NCBL Report at 28-33. Nevertheless, as I have shown elsewhere, a simple nonlinear model that considers some crude statistics on the college attended can boost the correlation of FYA with grades by 25 percent. *See* Kaye, *An "A" is an "A" is an "A": An Exploratory Analysis of a New Method of Adjusting Undergraduate Grades for Law School Admissions Purposes*, 31 J. LEGAL EDUC. (forthcoming 1980).

It is important to observe that the entire enterprise of comparing the LSAT's correlation with FYA to that of undergraduate grades may be misdirected. For operational purposes, the more relevant question is whether using the LSAT *in combination with* grades to predict performance is superior to using one of these factors alone. The answer to this question is clearly affirmative. The 1976 Carnegie study mentioned above, *see* p. 447 *supra*, for example, reports that the median validity when both LSAT scores and grades are used in a "multiple" regression equation is .45, compared with .36 for the LSAT alone and .25 for grades alone. Willingham & Breland, *supra*, at 237. This finding is one of the reasons LSAC cautions schools against using a specific LSAT score as a minimum for consideration for admission. *See* Cautionary Policies, *supra* note 17.

90. As I observed above, *see* note 86 *supra*, accuracy varies from school to school. Although it is sometimes argued that r^2 understates or misstates the accuracy of predictions, *see, e.g.*, Dawes, *Graduate Admissions Variables and Future Success*, 187 SCI. 721 (1975); Wolins, *On Squaring Validity Coefficients*, 37 EDUC. & PSYCH. MEASUREMENT 373, 374 (1977), the values of r^2 for the LSAT at most law schools are hardly overwhelming. Given the limited power of the LSAT to predict first-year grades, the present system of reporting LSAT scores to three digits must be questioned. Analysis of data on the 1978-79 first-year class at Arizona State University College of Law shows that compressing the scale by a factor of 40, that is, treating scores of 200-239 as a 1, scores of 240-279 as a 2, and so on, does not reduce the correlation between a weighted sum of LSAT scores and undergraduate grade point average (UGPA) on the one hand, and FYA on the other. Thus, the practice of three-digit reporting may encourage admissions personnel to attach importance to statistically insignificant differences in LSAT scores.

But is the LSAT equally accurate in predicting the grades of blacks, Hispanics, and whites? Although most courts have assumed in Title VII cases that an inference of discrimination based on disparate impact is refuted by proof of predictive validity,[91] a more refined analysis is plainly necessary. The general finding that a test beats rolling dice or predicting by the mean still may mask invidious differences in the predictions for various subgroups. For example, a test that consistently underpredicts the performance of blacks while accurately forecasting the achievements of all other subgroups will have a high r^2 if the degree of underprediction is not very large or if blacks make up a small percentage of the total group. Yet such a test unquestionably discriminates against blacks.

The statistical procedure for investigating the possibility of this type of bias in an apparently valid test is a simple extension of the least squares method. One merely calculates the regression lines for the subgroups in question and observes whether there are any large differences in the slopes, location, and fit of the various lines.[92] Every study of this nature has found that the LSAT does not systematically underpredict the initial academic performance of blacks or Hispanics.[93] For these subgroups, the LSAT is about as good a predictor of first-year law-school grades as it is for whites.[94] In sum, the principal charges against the LSAT—that it discriminates on the basis of race and that it has virtually no power to pre-

91. A search of the LEXIS library of federal cases reveals that very few of the Title VII cases that have referred to test validity have mentioned either underprediction or differential validity. In contrast, the Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. §§1607.7, 1607.14 (1979), recognize the importance of analysis of predictive validity by subgroups.

92. For examples of this method, see Hunter, Schmidt, & Hunter, *Differential Validity of Employment Tests by Race: A Comprehensive Review and Analysis*, 86 PSYCH. BULL. 721 (1979) (finding that most employment tests are equally accurate in predicting job performance of blacks and whites); Linn, *Test Bias and the Prediction of Grades in Law School*, 27 J. LEGAL EDUC. 293 (1975). A related, statistically superior technique involves inserting "indicator" variables for race into the equation for FYA. *See* Fisher, *Multiple Regression in Legal Proceedings*, 80 COLUM. L. REV. 702, 722 (1980).

93. *E.g.*, Powers, *Comparing Predictions of Law School Performance for Black, Chicano, and White Law Students*, in 3 REPORTS OF LSAC SPONSORED RESEARCH, 1975-1977, at 721 (1977) (Report No. LSAC-77-3); Pitcher, *Subgroups Validity Study*, in 3 REPORTS OF LSAC SPONSORED RESEARCH, 1975-1977, at 413 (1977) (Report No. LSAC-76-6); Schrader & Pitcher, *Prediction of Law School Grades for Mexican American and Black American Law Students*, in 2 REPORTS OF LSAC SPONSORED RESEARCH, 1970-74, at 715 (1976) (Report No. LSAC-74-8). These studies generally consider predictions made by the LSAT singly and in combination with UGPA.

94. In contrast, UPGA has been found to be consistently less valid for predicting law-school FYA of minority students. Powers, *supra* note 93, at 747. The NCBL Report ignores this finding in arguing that UPGA is a fairer and more valid predictor of FYA for blacks.

dict grades in the first year of law school—partake more of rhetoric than of reality.

Conclusion

This analysis does not mean that the LSAT is the be-all and end-all in law school admissions. I have argued that, to the extent one is interested in finding out how a prospective student will do in his first-year courses, the LSAT supplies a statistically useful bit of information, and that this information has roughly the same probative force for minority and majority candidates.[95] Nonetheless, admissions personnel should be interested in other information as well.[96] No one has an *a priori* right to attend law school simply be-

95. To put it another way, the LSAT discriminates against blacks and Hispanics in the same way that law school grading systems do. *Cf.* note 78 *supra* (discussing reasons for using FYA as criterion for predictive validity). Some psychometricians contend that for a test to be used fairly, group membership as well as the individual likelihood of success must be considered in defining cutoff lines. *See, e.g.*, Cole, *Bias in Selection*, 10 J. EDUC. MEASUREMENT 237 (1973); Thorndike, *Concepts of Culture Fairness*, 8 J. EDUC. MEASUREMENT 63 (1971). The three studies reviewed here refer to this literature in order to cast doubt on the generally accepted regression model of fairness. Indeed, the NCBL Report, in a novel extension of the Thorndike and Cole models, recommends the deliberate boosting of minority applicants' scores.

None of the reports, however, recognizes the logical inconsistencies in the Cole and Thorndike definitions of test fairness. *See* Linn, *supra* note 92 (noting such inconsistencies); Novick & Ellis, *Equal Opportunity in Educational and Employment Selection*, 32 AM. PSYCH. 306 (1977) (same). Moreover, it is not clear that adopting an alternative quantitative definition would necessarily open the law school doors to blacks and Hispanics. A study of ACT scores, high-school rank, and first- and second-year grades at one university, for instance, found that substitution of the Thorndike definition for the conventional regression model of test bias would have "changed the results for minorities very little," and would have required that admissions standards be *raised* for blacks and lowered for Jews. Silverman, Barton, & Lyon, *Minority Group Status and Bias in College Admissions Criteria*, 36 EDUC. & PSYCH. MEASUREMENT 401, 405 (1976). In short, it seems hard to avoid the conclusion that a test that gives equally accurate predictions for minority and majority applicants does exactly what it should be designed to do. What use should be made of these predictions is another matter, and it is unfortunate that most of the alternative definitions of test fairness conflate the two issues.

96. *See* ABA SECTION OF LEGAL EDUCATION AND ADMISSIONS TO THE BAR, REPORT AND RECOMMENDATIONS OF THE TASK FORCE ON LAWYER COMPETENCY: THE ROLE OF THE LAW SCHOOLS 3 (1979) (Recommendation 1).

The findings of an ETS study are provocative. Questionnaire responses of college seniors who had taken the LSAT revealed that higher scoring students tended to be academically superior, but to have less confidence in their ability to relate to others on an individual basis, and to place less importance on security, salary, working with people, and making a contribution to knowledge in choosing law as a career. Baird, *Biographical and Educational Correlates of Graduate and Professional School Admissions Test Scores*, 36 EDUC. & PSYCH. MEASUREMENT 415, 419 (1976). In light of the mild negative correlation between the LSAT and those personal traits (some which may be important in a lawyer), and in view of the questionable ability of the LSAT to make meaningful distinctions at the upper end of the score scale, one admissions strategy worth considering would be to use the LSAT in conjunction with UGPA and related factors to screen out applicants who present a serious academic

cause he or she is unusually clever or articulate.[97] In particular, blacks and Hispanics are being admitted to law schools under legitimate affirmative-action programs that do not, ostrich-like, hide from the fact of disproportionately low LSAT scores among those groups.[98]

What is disturbing, then, about the recent spate of law-school admissions exposes, probes, and studies is not the concern with enhancing minority admissions. Rather, it is the meritocratic assumption—shared by all three reports under review—that admission to law school is a reward for high undergraduate grades. It is the common theme, sounded most stridently in the *Nader* and NCBL reports, that because use of the LSAT may obstruct efforts to increase minority admissions, the test should be scrapped. But experience has shown that, for the present, only overtly race-conscious programs that guarantee seats for minority candidates will do much to achieve integration in law schools and in the practicing bar.[99] Moreover, it is doubtful that discarding the LSAT would substantially aid minority enrollments.[100] At the same time,

risk, then to proceed with the selection on the basis of non-academic characteristics. Another strategy that may prove more appealing as the number of applications to law school subsides would be to minimize mistaken predictions of academic success by admitting all minimally qualified applicants with the clear understanding that many, if not most, of those admitted will not meet the school's announced standards for retention and graduation.

97. R. DWORKIN, TAKING RIGHTS SERIOUSLY 225 (1977).

98. *See* Evans, *supra* note 32, at 634-36.

99. *See id.* (analyzing the impact on law school enrollments of race-neutral selection methods that incorporate consideration of LSAT).

100. All three reports insist that the LSAT has more disparate impact on minority admissions that do undergraduate grades. The *Nader Report*, for example, laments that "[t]he lower LSAT scores generally received by black and Chicano students . . . put them in the position of having to earn higher college grades than their white counterparts for an equal chance of admission," NADER REPORT at 225, and the NCBL Report adopts the same perspective, *see* NCBL Report at 34-38. How this statement supports the assertion that the LSAT is *more* discriminatory than undergraduate grades is a mystery. After all, one could also say that the lower grades generally obtained by minority applicants, *see* note 72 *supra*, force them to earn higher LSAT scores than white students.

The NCBL Report also maintains that admitting minority students "on the basis of UGPA *alone* is considerably more favorable" than admitting "under a race-blind system using LSAT *and* UGPA." *Id.* at 37. The statistical support provided for this conclusion, however, is sheer sophistry. The report compares the *hypothetical* rates at which minorities would have been accepted in 1976 had admissions been based on a linear combination of LSAT and UGPA to the *actual* rate at which minority students with various UGPAs were accepted. Inasmuch as affirmative action programs were in place to aid minority students with low UGPAs in 1976, this comparison proves very little. The better comparison is between race-blind admissions based on LSAT alone, UGPA alone, and UGPA and LSAT combined. The information presented in Evans, *supra* note 32, at 604, on the distributions among ethnic groups of LSAT scores and the UGPAs in the class entering law schools in 1976 is not sufficiently detailed to allow this comparison. Nevertheless, by performing some linear interpolations to subdivide the categories used in that report, it appears that if

it is plain that ignoring LSAT scores would not help many law schools discern which applicants—of whatever race or ethnic background—are academically qualified. It seems preferable to use the LSAT and related data to make the best possible judgment as to academic risk, and to select an appropriate mix of students with

the available places had been allocated by LSAT alone, 13 percent of the black candidates, 23 percent of the Hispanic applicants, and 61 percent of the white candidates would have been accepted. Almost the same pattern would have resulted if acceptances had been distributed by LSAT combined with UGPA. The hypothetical acceptance rates are then 11 percent, 23 percent, and 61 percent, respectively. Under an admissions process that considers UGPA alone, there would be a distinct improvement for minorities. This system would produce distinct acceptance rates of 30 percent, 41 percent, and 59 percent, respectively. These calculations tend to substantiate the view that selection by UGPA has a less severe adverse impact on blacks and Hispanics than does selection by LSAT.

The true picture, however, may not be adequately depicted by these simple calculations. The derivation of these figures by linear interpolations is inexact, because the actual distributions are obviously nonlinear. In addition, admission rates based on a system that looked to undergraduate grades alone would be complicated by consideration of the colleges attended and the fields of study pursued by minority and majority candidates. Finally, even if the UGPA figures are uncritically accepted, they still represent admission rates for blacks and Hispanics that are *below* the status quo. Despite the heavy weighting of the LSAT that the *Nader*, NCBL, and *MALDEF* reports decry, 39 percent of the black applicants, 47 percent of the Hispanic candidates, and 59 percent of the white applicants were admitted to law schools in 1976. *Id.* at 604.

The *MALDEF Report* relies on its own study of three San Francisco Bay Area law schools to show that UGPA has a less disparate impact. It finds that the mean rankings of minority groups are higher when the relative weight given to the LSAT is decreased. MALDEF REPORT at 60. The report does not state in absolute terms how much improvement in the rankings is obtained. Neither does it state what effect the shift in rankings would have had on actual offers of admissions. Indeed, the *MALDEF* study looked only to shifts among students *already accepted* to law school, making it impossible for the researchers to gauge directly the effects on minority *applicants*.

Because it is offers of admission, not rankings of students who are admitted, that is of importance to persons who want to enter law school, I used the data on the most recent group of applicants to the Arizona State University College of Law to gain some insight into this question. The results were consistent with the *MALDEF Report*'s findings. Selection based solely on the LSAT would have admitted only one minority student, a Native American. Rankings on the basis of the expression LSAT + 120 × UGPA (the optimal linear combination of LSAT and UGPA in the least squared sense for Arizona State) would have admitted two Native Americans and three Hispanics. Using UGPA alone would have admitted one Native American, six Hispanics, and three blacks.

This limited demonstration of the relatively unfavorable impact of the LSAT is subject to many of the same caveats as is the previous analysis of the national-applicant data. For example, if attention were paid to such factors as the field of study and the college attended, the gains associated with undergraduate grades might disappear. *But see* NCBL Report at 38-43 (arguing that proper consideration of such factors would not disadvantage minorities). In addition, compared to the existing admissions process at Arizona State, which uses the weighted combination of LSAT and UGPA along with other factors, admission by UGPA alone would have proved disastrous for minority admissions. Under the existing program, 28 Hispanics, 10 blacks, and 10 Native Americans were admitted. To be sure, the critics of the LSAT might not wish to see race-blind adherence to UGPA, but it is not obvious that admissions by undergraduate grades and other factors would produce any more minority enrollments than a system based on undergraduate grades, LSAT, and other factors.

due regard to other considerations, including minority status.[101] Fairness in selection and accuracy in prediction are complementary virtues, not implacable antagonists.

101. The *MALDEF Report* offers some suggestions as to how this might be done. MALDEF REPORT at 39-58. It also remarks on the important role of supportive programs. *Id.* at 63-78.

Doctor-Patient Dialogue: A Second Opinion on Talk Therapy Through Law

Taking Care of Strangers: The Rule of Law in Doctor-Patient Relations. *By* Robert A. Burt. *New York: The Free Press*, 1979. Pp. vii, 200. $15.95

David B. Wexler†

I

Robert Burt's *Taking Care of Strangers*[1] is both fascinating to read and frustrating to review. Drawing heavily on psychoanalytic theory, the book proposes a legal mechanism for designing an informed consent doctrine of doctor-patient relations that is far more fluid, dynamic, and interactional than the current law.

The book makes fascinating reading (my first reading of it made almost enjoyable a six hour delay at National Airport) because, in attempting to construct a new notion of informed consent, Professor Burt makes an enormous number of creative, insightful, and thought-provoking comments on an extremely varied range of topics, including the impact of the abolition of civil commitment laws,[2] the incest taboo,[3] the problem of child abuse,[4] the phenomenon of irrational and infantile thinking in adults,[5] the experiments of Stanley Milgram on the propensity to obey malevolent scientific authority,[6] the placebo effect,[7] and the development of the informed consent doctrine.[8]

The book makes frustrating reviewing because nowhere does Burt set forth a step-by-step guide to what he is proposing, why he is proposing it, or how his proposal might actually be implemented.

† Professor of Law, University of Arizona.

1. R. BURT, TAKING CARE OF STRANGERS: THE RULE OF LAW IN DOCTOR-PATIENT RELATIONS (1979) [hereinafter cited by page number only].
2. *See* pp. 32-34, 44.
3. *See* pp. 56-60.
4. *See* pp. 61-65.
5. *See* pp. 50-54.
6. *See* pp. 72-91 (discussing S. MILGRAM, OBEDIENCE TO AUTHORITY (1974)).
7. *See* pp. 104-06.
8. *See* pp. 102-03.

Due to this shortcoming, the major difficulty I have with the work, the book is particularly prone to idiosyncratic interpretation, or to outright misinterpretation, by readers. How difficult the book is to interpret becomes especially evident when the reader turns reviewer and attempts to summarize the book, and to piece portions of it together, in order to launch a review. My interpretation follows.

All adults, Burt tells us, whether they be doctors, patients, commitment officials, or ordinary individuals, retain alongside their rational "secondary process" thinking an irrational "primary process" thinking characteristic of infants.[9] Infantile thinking is full of paradoxical and unclear distinctions: distinctions between the boundaries of self and others, between feelings of impotence and omnipotence, and between feeling pain and experiencing the pleasure that nurturing parents insure will quickly follow the pain.[10]

Primary process thinking in adults is particularly noticeable in abusive parents. Such parents, Burt claims, have difficulty distinguishing their own pain from their child's pain. They may inflict pain on the child in order to reexperience the pleasure of feeling pain. At the same time, the child's pain reminds those parents of their own painful childhood, and the child's ability to trigger such unpleasant memories makes the child seem to have enormous power over them. Those parents may abuse the child in order to silence him and obliterate that memory. Abusive conduct also restores the parents' feeling that they are omnipotent in their relation to the child. Abusing parents are distinguishable from nurturing ones particularly in their inability to tolerate their incapacity to separate their own pain from their child's.[11]

Although primary process thinking is especially evident in abusive parents, it may surface in any adult, including doctors and commitment officials, especially when those officials deal with individuals who appear bizarre, alien, and constantly pained.[12] Burt provides examples of such abuse-prone persons: Mr. G, a blinded, severely burned victim who resisted treatment,[13] and Mrs. Lake, a disoriented, "wandering" woman who was committed to a mental institution in the District of Columbia.[14] Benevolent and abusive motives fuse when we "care for" such persons, for "the pessimism

9. Pp. 51-54.
10. P. 50.
11. *See* pp. 61-65.
12. *See* p. 65.
13. Pp. 1-21; *see* note 36 *infra*.
14. Pp. 22-45; *see* Lake v. Cameron, 364 F.2d 657 (D.C. Cir. 1966).

that peculiarly afflicts the abusing parents in their perceptions of their children, and that triggers their abusive conduct, is everyone's common perception of Mr. G and Mrs. Lake."[15]

Burt believes that in typical, healthy interpersonal relations, no person conceives of himself, vis-a-vis another person, as *wholly* choice-making or *wholly* choiceless.[16] Instead, "each party typically conceives himself alternatingly through one mode and the other."[17] A patient like Mr. G or Mrs. Lake, however, provokes us to adopt an unhealthy, rigid distinction between ourself and that person, between being wholly choiceless and being wholly choice-making.[18] Invoking the vagaries of psychoanalytic theory, Burt explains the danger of such a distinction:

> Whenever anyone seeks to conceive himself as either a wholly choice-making or wholly choiceless individual regarding another person, that aspiration reflects the individual's uncertainty regarding the boundaries between his "self" and the other's and an unwillingness consciously to acknowledge his uncertainty. This unwillingness finds expression in an intrapsychic attempt to conceive both people as one, thereby attempting the *tour de force* of suppressing the belief that the two people should be conceived as separate entities. Unless the attempt to maintain this univalent conception is interrupted, the individual is led toward action to destroy the physical existence of that other person and/or himself in order to keep his intrapsychic grasp on the constructed "self" which depends on the obliteration of the "other's" alien separateness.[19]

In Burt's view, the law should help us avoid the destructive consequences that flow from our temptation to view abuse-prone people like Mr. G or Mrs. Lake as either wholly choice-making (Mr. G wants to die, so the doctors have no say; Mrs. Lake is free to lead her life as she wishes, so there is nothing the commitment officials can do to prevent her from wandering) or as wholly choiceless (Mr. G's desire to die is irrational, so we will treat him despite his expressed wishes; Mrs. Lake is senile and confused, so she needs long-term institutionalization). Rather, the law should structure

15. P. 65. Some have even speculated that certain types of children may produce abusive behavior in parents, rather than vice versa. Skolnick, *The Myth of the Vulnerable Child*, PSYCH. TODAY, Feb. 1978, at 56, 58.

16. P. 119.

17. *Id.*

18. P. 25.

19. Pp. 118-19.

medical and psychiatric decisionmaking to serve as quasi-psychoanalytic sessions, fostering oscillation in self-conceptions and preventing patients and other participants from fixing on univalent conceptions.[20]

To achieve this healthy oscillation, the law should encourage ongoing communication and negotiation between doctor and patient.[21] Burt proposes that the law encourage increased communication by refusing to settle doctor-patient debates conclusively at the pre-treatment stage.[22] Instead, doctor and patient should confront each other and seek to come to treatment terms knowing that the law may review the scenario after the fact.[23] The possibility of post-treatment review and an uncertainty about civil and criminal liability should, Burt believes, provide a framework for a conversation that will make less likely the adoption of destructive univalent conceptions.[24]

To limit the law's role to predominantly after-the-fact review, Burt proposes that civil commitment laws be reformed to allow for only short-term commitments, during which doctors might be limited to those treatment modalities "clearly intended and necessarily limited to fostering conversation"[25] between doctor and patient. In the case of Mrs. Lake, then, officials could confine her only for a short stay, at the expiration of which she would have to be released "unless [the doctors] could persuade her to remain or unless they were willing to run risks of [civil or criminal] liability for retaining her longer."[26]

In Mr. G's case, the doctors could either respect Mr. G's wishes not to be given skin grafts and painful chemical immersions, or they could confine and treat him over his objections, hoping that no civil or criminal action would result.[27] A pre-treatment resolution of the rights and liabilities of the doctors and Mr. G would be unavailable unless the doctors were able to establish Mr. G's resistance to treatment as irrational, as temporary mental incompetence.[28] Even in the rather unlikely event that the doctors would

20. P. 119.
21. *See* pp. 120-21.
22. P. 121.
23. Pp. 126-27, 137.
24. *See* p. 137.
25. P. 129. Burt does not indicate whether his proposal would apply to guardianship laws, but presumably it would.
26. P. 141.
27. Pp. 131-33.
28. Mr. G was severely burned as he turned his automobile ignition switch. His father

seek or obtain a civil commitment or guardianship order in Mr. G's case, any order they obtained would, under Burt's scheme, be limited in time and in the treatment it permitted, and would be designed to foster conversation.[29] The possibility of post-treatment review would constitute the major legal force shaping the doctors' actions regarding Mr. G.

For dealing with individuals like Mr. G and Mrs. Lake, then, Burt proposes severely curtailing both the length of time for which such individuals may be committed and the types of treatment authorized during commitment. Burt also applies his theories to "silent" patients: patients who are comatose, as in the celebrated case of Karen Quinlan,[30] or severely mentally retarded, as in the case of Joseph Saikewicz, a Massachusetts man suffering from leukemia whose life could have been prolonged by chemotherapy treatments.[31] In those kinds of cases, Burt would make unavailable declaratory judgments on the propriety of administering or withholding treatment. Burt believes that, facing the possibility of later civil or criminal liability, the doctors and family members would converse and agonize until a consensus emerges from a thorough airing of the matter.[32]

Burt's thesis raises two principal questions: Will the proposal in fact increase communication, and will increased communication actually improve decisionmaking?

died in the accident. There is some psychiatric evidence that Mr. G's guilt over the incident caused him to believe he did not "deserve" to live. P. 10. Additionally, there is some evidence that Mr. G wanted to die principally because he believed others, including family members, wanted him to die. P. 11.

29. P. 129. Particularly because Mr. G reported the chemical immersions as "*very* painful," p. 180, I am unable to understand how Burt applies his own commitment and treatment proposals to the facts of Mr. G's case:

> Mr. G's situation can illustrate my vision for reformation of the law. If the civil commitment law provided that a finding of mental illness justified overriding Mr. G's protests only for a limited time period (say, thirty days) and *only with treatment modalities that were clearly intended and necessarily limited to fostering conversation* between him and his physicians, that would mean Mr. G could be forced to continue the chemical immersions for thirty days and to meet with the psychiatrist and other physicians, but that no law would force him to acquiesce in the operation on his hands. *The immersions would not be authorized because they were necessary to keep him alive and thus available for conversation, but rather because the immersions themselves were not central to* and did not end the dispute between him and his physicians. A time-limited continuation of the immersions could be seen by both parties as temporizing and thus as an inducement toward, rather than the end of, conversation.

P. 129 (emphasis added).

30. *In re* Quinlan, 70 N.J. 10, 355 A.2d 647 (1976).

31. Superintendent of Belchertown State School v. Saikewicz, 373 Mass. 728, 370 N.E.2d 417 (1977).

32. Pp. 165-67. Burt's emphasis on agonizing is explicit. Pp. 127, 165, 167.

II

Burt recognizes that "[i]t may be that no physician today would knowingly risk even a dollar fine or damages payment for holding Mr. G against his expressed will."[33] He also recognizes that unless and until physicians "come to believe that only a limited or bearable adversity might result if they disregarded Mr. G's claim for release,"[34] his proposal would not lead to increased communication, but would lead instead to the prompt termination of communication *and* treatment.[35] If that is the result, persons such as Mr. G would, curiously, receive *less* treatment under Burt's model than they would under laws authorizing pre-treatment legal intervention.[36]

33. Pp. 132-33.
34. P. 133.
35. *See id.*
36. Though it is not an explicit component of Burt's thesis, a value preference for treatment over non-treatment clearly runs through his book. *See, e.g.*, pp. 165, 167 (emphasizing "prolonging treatment").

To begin to gain insight into the empirical question whether a person like Mr. G would be more likely to be treated over his objections under Burt's scheme or under a scheme authorizing pre-treatment adjudication, I administered a "Mr. G hypothetical" to fifty-three summer-session law students. Without being told that there were two versions of the problem, half of the students (Group A) were given a form asking them to decide the case in a pre-treatment context, and half (Group B) were given a form posing the matter in a post-treatment civil suit. The facts were stated as follows:

> Mr. G is a twenty-seven year old unmarried man who has always been very active in sports. After service in the air force, he joined his father's real estate firm. Soon thereafter, the two men went together to inspect some property. When their car failed to start, Mr. G's father lifted the hood to manipulate the carburetor and directed his son to start the ignition. Mr. G did so, and the car suddenly was enveloped by fire. Unknowingly, they had parked the car over a leaking gas main.
>
> Mr. G's father died on the way to the hospital. Mr. G had received severe burns over two-thirds of his body. He was not expected to survive, but he did, though blinded and terribly maimed. From the beginning of his hospitalization he suffered greatly. Doctors believed there was constant danger that fatal infections would enter his extensive open sores. Each day, to guard against this, he was immersed in a chemical-filled tank for excruciatingly painful treatments. Surgeons also performed several operations for skin grafts, for unsuccessful efforts to save his sight, and for restoring some movement to his limbs. Throughout this time, Mr. G had repeatedly expressed doubts about whether he wanted to live. Nine months after his accident, he adamantly refused further medical treatment. In an interview with Dr. White, a psychiatrist, Mr. G explained his refusal.
>
> Mr. G: What really, I guess, astounds me, I guess I'd say, is that in a country like this where freedom has been stressed so much and civil liberties, especially during the last few years, how a person can be made to stay under a doctor's care and be subjected to the painful treatment, such as the tankings which are *very* painful, against this person's wishes, especially if he has demonstrated the ability to reason.
> Dr. White: Even where discontinuing these tankings and the like would be a circumstance that would mean that infections would set in that would end your life.
> Mr. G: Even then. The way I see it, who is a doctor to decide whether a person lives

Deinstitutionalized mentally disabled persons, who are mounting in numbers,[37] run a serious risk of being denied important medical attention if Burt's scheme backfires. That result seems likely with such "silent" or "semi-silent" patients. When a mentally retarded person resides in the community, the person has no clear "right" to be treated by a given doctor, and the doctor has no legal "obligation" to treat such a person. If an arguably incompetent person requires, or simply requests, a risky medical or surgical procedure,

> or dies just because someone has put him under your care as a doctor? As long as the patient is willing to be treated, I certainly think that a doctor should do everything they should, they could.
> Dr. White: But you feel that you properly should have the legal right to say, "No, I do not want to be treated."
> Mr. G: Yes, I don't see how anyone else could possibly have this right, justifiably have this right. That's what, like I said, really astounds me.
> [P. 180.]

In Dr. White's view, Mr. G was in a confused state. Guilty over his father's death, which came about from Mr. G's turning the ignition switch, Mr. G was struggling psychologically with the question whether he himself "deserved" to live. Further, in Dr. White's view, Mr. G was choosing death because of a nagging feeling that others—such as his next of kin—"really" wished he would die.

Group A: Sharing Dr. White's perspective, Dr. Brown, the physician overseeing Mr. G's treatment, wished to treat Mr. G despite G's expressed wishes. Dr. Brown goes to court seeking authorization to treat Mr. G.

Group B: Sharing Dr. White's perspective, Dr. Brown, the physician overseeing Mr. G's treatment, wished to treat Mr. G despite G's expressed wishes. Under local law, there is no procedure for seeking, in advance of treatment, a court ruling regarding the propriety of proposed treatment. Dr. Brown treated Mr. G, and Mr. G then filed suit against Dr. Brown.

Assume that you are the decisionmaker, that the governing law is unclear, and that your decision should be based simply on your intuitive inclinations.

In each group (one of 26 members, the other of 27), only eight respondents (fewer than one-third) ruled in favor of the doctor. Because of the limited sample size, I did not attempt to ascertain how comfortable with their decisions the decisionmakers were, nor did I ascertain the amount of damages they would have awarded in the post-treatment suit. I did collect comments from several of the participants. One student who had answered a post-treatment form remarked upon turning in her form: "I sure wish I could have decided this case *before* treatment. Then I would surely have ruled for G. As it is, I ruled for G anyway, but it was difficult to do so, and I surely wouldn't have given him much in the way of damages." Another student, who ruled for the doctor in the post-treatment context, commented that she would not have if the doctor had had the opportunity to seek pre-treatment review and had failed to do so. *But cf. In re* Spring, 405 N.E.2d 115, 122 (Mass. 1980) (negligence should not be based solely on failure to seek prior court approval if such approval would have been given).

Upon learning that two out of three courts would rule against the treating physician in the Mr. G post-treatment hypothetical, it is likely that doctors would refrain from treating similarly situated objecting patients. But under a legal system in which there seems to be a one in three chance of obtaining a pre-treatment commitment or guardianship order, one-third of the similarly situated patients may well be subjected to treatment.

37. Task Panel on Legal and Ethical Issues, President's Comm'n on Mental Health, *Mental Health and Human Rights: Report of the Task Panel on Legal and Ethical Issues*, 20 ARIZ. L. REV. 49, 72 (1978).

doctors, if they are unable to gain a pre-treatment determination of competency, are likely to refuse the treatment. The fear of later liability premised on the patient's inability to give competent consent, particularly in those situations in which there are significant inherent risks, such as paralysis in orthopedic surgery, will discourage doctors from treating patients arguably incapable of consenting. Doctors unable to obtain a pre-treatment adjudication of patient competence or incompetence will be extraordinarily reluctant to perform irreversible procedures, such as sterilizations and abortions, on mentally retarded persons.[38]

Instead of increasing communication, Burt's proposal may thus, in certain circumstances, simply terminate the conversation between doctor and patient. Even if the quantity of "talking" by doctors increases, however, it is not clear that genuine two-way "communication" will be advanced. In Mrs. Lakes' case, suppose that the doctors, using skills of persuasion and perhaps "drugs . . . to calm her in conjunction with attempts to talk to her,"[39] "talked her into" remaining in the hospital and submitting to tranquilizing medication. If a major concern of the doctors is whether Mrs. Lake "would later sue them,"[40] the doctors would be able and likely to confine her with impunity so long as they keep her sedate, appreciative, and non-litigious.

My reading of *Taking Care of Strangers* suggests that this troubling scenario is quite possible should Burt's proposal be implemented. Yet, it is difficult to believe that Burt would support such a result, especially in light of his concern that, under existing law, a person like Mrs. Lake might "be led explicitly to consent to

38. *See* Ruby v. Massey, 452 F. Supp. 361, 364 (D. Conn. 1978) (in absence of procedure for ascertaining legal effectiveness of patient consent, doctors and hospitals refused to perform sterilizations on mentally retarded persons); Petro v. McCullough, 385 N.E.2d 1195 (Ind. Ct. App. 1979) (suit against doctor, by mildly retarded and previously hospitalized patient, claiming tubal ligation was performed with consent patient was incompetent to give).

If Burt's proposal has the effect of making pre-treatment adjudication unavailable in such instances, and if the inability of physicians to obtain pre-treatment guidance virtually forecloses the possibility of mentally retarded persons securing constitutionally protected sterilizations and abortions, Burt's scheme may be deemed unconstitutional. In *Ruby*, the statutory silence and ambiguity regarding legal procedures for sterilizing noninstitutionalized retarded persons led doctors and hospitals to refuse to perform such procedures on mentally retarded persons residing in the community. 452 F. Supp. at 364 & n.13. Plaintiffs challenged the de facto denial of the right to sterilization on grounds of privacy, due process, and equal protection. *Id.* at 363. The court considered only the equal protection claim and found it meritorious. *Id* at 369.

39. P. 141.

40. *Id.*

lifetime hospitalization on the back ward of a mental hospital,"[41] his psychodynamic analysis concluding that even the abolition of civil commitment laws would probably not alter that destructive dynamic,[42] and his penetrating critique of the ethics of Stanley Milgram's experiments on obedience to malevolent scientific authority.[43] In the Milgram experiments, subjects complied with a scientist's directives and administered to a third person what the subjects thought were painful electric shocks. Burt provides a convincing argument that subjects who were led by the mantle of science to succumb to these so-called malevolent wishes of an experimenter could be led, by the very same process, to tell the experimenters that they were not emotionally pained by having participated in the indisputably stressful experiment.[44] Why would not any acquiescence to hospitalization by a Mrs. Lake also constitute mere "compliance with the [scientific doctor's] obvious wishes"?[45] The same analysis applies to the case of Mr. G. He ultimately consented to medical treatment and now, despite a stormy period immediately after he first left the hospital, is apparently doing well.[46] Why should we conclude, as Burt does, that "each day that [Mr. G] lives can be more convincingly characterized as his choice for himself . . ."?[47] Why should we not conclude that Mr. G's ultimate acquiescence in the treatment, his failure to sue the doctors, his belated appreciativeness, and his decision not to take his life are

41. P. 38. *See also* p. 41. For Burt's critique of the "consensual format," see p. 39.

42. *See* pp. 32-34, 41-45. Burt believes that since "[t]he stress created in others by Mrs. Lake's confused appearance was not an artifact of the civil commitment laws," p. 33, but was attributable instead to Mrs. Lake's confusion about the conceptual boundaries between herself and others, the abolition of civil commitment laws would not significantly alter the destructive forces that Mrs. Lake unleases in others. Instead, she is likely to be pressured into consenting to "voluntary" hospitalization, p. 41, "pressed by implicit and explicit social pressures to ask for her self-obliteration," p. 34.

As noted in the text, however, I fail to see how abolition differs from doctor-patient relations guided only by possible after-the-fact judicial review. The abolition of civil commitment laws would not, of course, undo existing after-the-fact civil and criminal penalties for wrongful or coerced confinement. Thus, if civil commitment laws were abolished *either* doctors would refrain from abusing Mrs. Lake for fear that they may later be sued or prosecuted for wrongful imprisonment, in which case Burt is wrong in his assessment of the impact of abolition, *or* doctors would abuse Mrs. Lake, in which case Burt is wrong in his assessment of the salutary effects of after-the-fact review!

43. *See* pp. 86-89.

44. "Their assurances that they felt no pain were, it seems to me, as much compliance with the experimenter's obvious wishes as their earlier compliance with the directive [to administer the electric shock]." Pp. 86-87.

45. P. 87.

46. P. 122.

47. *Id.*

all nothing more than a patient-subject's compliance with the wishes of his medically clothed experimenter?

III

We have seen, then, that Burt's legal system, rather than increasing the quantity and quality of communication, may cut off conversation or induce one-sided, coercive conversation. Even if an increase in full-fledged communication were to be a result of Burt's proposal, however, there is some question whether the increased communication would actually improve the decisionmaking process. Not surprisingly, we feel more confident in our decisions when they are reached after spending significant amounts of energy and effort poring over details. Ordinarily, of course, there is a positive payoff to that course of action: decisions are generally likely to be "better" if they are made after careful consideration, effort, and even agony.

Sometimes, however, particularly when we deal with important matters having no easy answers, we delude ourselves into believing that the process of agonizing will somehow improve our decisions. But agonizing can in some situations be costly and even counterproductive. Lengthy medical school admission interviews have been criticized because they lack predictive power and may even lower the quality of the admission decision.[48] One important study on the related topic of predicting success in college revealed that "[a]ll six of the [predicting] psychologists were able to make better predictions with [four] predictors than they were with [twenty-two] predictors."[49] Another study, dealing with the processing of mentally disordered sex offenders (MDSOs), found that although participating judges and psychiatrists may well *believe* that their effort and expertise is important in determining the outcome of cases, whether or not a person is found to be an MDSO depends on the defendant's prior sex-related criminal record, a variable that *precedes* judicial and psychiatric consideration of the particular case.[50]

48. La Brecque, *On Making Sounder Judgments*, PSYCH. TODAY, June 1980, at 33, 42.

49. Bartlett & Green, *Clinical Prediction: Does One Sometimes Know Too Much?* 13 J. COUNSELING PSYCH. 267, 268 (1966). Sometimes, however, by "diluting" what would have otherwise been too extreme a prediction, the exposure of clinicians to technically irrelevant factors serves to *improve* prediction. R. NISBETT & L. ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT 154-56 (1980).

50. Konecni, Mulcahy, & Ebbesen, *Prison or Mental Hospital: Factors Affecting the Processing of Persons Suspected of Being "Mentally Disordered Sex Offenders,"* in NEW DIRECTIONS IN PSYCHOLEGAL RESEARCH 87, 112-15 (P. Lipsitt & B. Sales eds. 1980). The authors have also

In some instances, communicating and deliberating more can be simply wasteful. Costly decisionmaking strategies are, of course, least appropriate for "decisions that, although of critical importance, have outcomes that are unknowable in principle or that are presently indistinguishable in terms of their overall costs and benefits."[51]

Burt can be read to suggest that the issues that concern him are often virtually unknowable.[52] Is Mr. G's wish to die really his own wish, or is it his attempt to implement what he perceives to be the wishes of others? If it is "really" his own wish, is it based on the reality of his present condition and his potential for future improvement, or is it based on the fact that he feels guilty for surviving the holocaust that killed his father? If the latter is the case, should that fact affect our decision on whether to respect his wish to die? Is Mr. G making a decision during a momentary state of balance between the alternative concepts of himself as a choice-maker or a choice-taker, or is he making the decision while holding to an ideology of self that lies on either end of the spectrum? When Karen Quinlan's father seeks to ascertain whether Karen would want to die, can he escape a psychological *renvoi*, insightfully illuminated by Burt,[53] asking himself whether Karen would want to know how *he* felt?

It may be that when we are called upon to make important decisions but are unsure of what substantive standard to employ, or are unsure of what evidence must exist to ascertain whether an agreed-upon substantive standard has been met, our tendency is to compensate by resorting to increased "process."[54] In those situations, however, such "process" is inefficient.

shown that criminal sentencing decisions are directly related to a defendant's prior bail status (jail, bail, or release on recognizance), that bail status is directly related to the prosecuting attorney's recommendation, and that the recommendation is directly related to the severity of the crime. *Id.* at 89; *see* Ebbesen & Konecni, *Decision Making and Information Integration in the Courts: The Setting of Bail*, 32 J. PERSONALITY & SOC. PSYCH. 805, 819-20 (1975). *See also* Shaprio & Clement, *Presentence of Information in Felony Cases in the Massachusetts Superior Court*, 10 SUFFOLK L. REV. 49, 58-59 (1975) (judges wanted more information in presentence reports and thought information would improve accuracy, but additional information is as likely to confuse as to help).

51. R. NISBETT & L. ROSS, *supra* note 49, at 279.

52. *See* pp. 15, 20, 37, 92, 121, 146.

53. P. 152.

54. It seems that "when the appropriate schema for analysis of a given situation is weak, then obvious and powerful theories used habitually in other domains will likely intrude." R. NISBETT & L. ROSS, *supra* note 49, at 136-37. Thus, people sometimes attempt to use "skills," such as thought, choice, effort, and involvement, in situations in which the outcome depends on "chance," and cannot be affected by the use of skills. Langer, *The Illusion of Control*, 32 J. PERSONALITY & SOC. PSYCH. 311, 322-23 (1975).

Should we subject patients, doctors, and families to what may be unnecessary agony? Karen Quinlan's devout parents were aided by a priest in reaching their decision to request termination of her life-support system.[55] Suppose such unequivocal support had not been forthcoming. Should the life-support system have been continued if Mr. Quinlan had failed to concur explicitly in a decision to terminate it?[56] If Mr. Quinlan had, under those circumstances, agreed reluctantly, might he not have later suffered psychologically for having agreed to a course of action not unequivocally endorsed by his church? Should the law force someone in that position to make such a choice?[57] Should not the role of the law be to reduce agony rather than to create it?[58]

Perhaps, however, increased agony generally produces psychological payoffs. The literature suggests that even if the answers to certain important questions are more or less unknowable, persons having a role in making important decisions may ultimately feel better and more confident about any decision made with their full involvement and after the airing of all factors that at least seem relevant to the matter's resolution.[59] Indeed, perhaps not only the

55. P. 162.

56. *See* p. 167 ("[t]reatment should be prolonged unless the face-to-face participants were unanimously agreed otherwise . . . .")

57. If, contrary to Burt's scheme, declaratory judgments were available, an individual would not be put in the position of either repressing his religious views or of expressing them knowing that they will then control the outcome. Declaratory relief makes possible an alternative in which the individual maintains his loyalty to his religious beliefs, and yet does not allow those beliefs to prevent him from receiving needed medical treatment. *See In re* President of Georgetown College, 331 F.2d 1000, 1006-07 (D.C. Cir.) (Wright, J., in chambers), *rehearing en banc denied*, 331 F.2d 1010 (D.C. Cir.), *cert. denied*, 377 U.S. 978 (1964) (Jehovah's Witness refused to consent to blood transfusion on religious grounds, but wanted to live and communicated to court that, if court ordered the transfusion, her religion would not be violated because decision would not be her responsibility).

58. *See* Wexler, *Criminal Commitment Contingency Structures*, in 1 PERSPECTIVES IN LAW AND PSYCHOLOGY: THE CRIMINAL JUSTICE SYSTEM 121, 128-33 (B. Sales ed. 1977) (proposing legal scheme for reducing unnecessary agony in decision to release institutionalized mental patients); Wexler, *Victimology and Mental Health Law: An Agenda*, 66 VA. L. REV. 681, 699-700 (1980) (proposing legal scheme allowing for face-saving by fathers in family therapy).

59. Decisionmakers seem to want large amounts of information, and feel more confident about decisions "based" on such information, even if the information seems technically not to have influenced the decision or if the information reduced the accuracy of the decision. *See* p. 467 & notes 48-50 *supra* (citing sources). Much evidence exists suggesting that we feel better about decisions over which we exercise control and even about decisions over which we *falsely believe* we exercise control. *See* Langer, *supra* note 54, at 323. *See generally* Langer & Rodin, *The Effects of Choice and Enhanced Personal Responsibility for the Aged: A Field Experiment in an Institutional Setting*, 34 J. PERSONALITY & SOC. PSYCH. 191 (1976) (giving nursing home residents increased responsibility improved their condition). *See also* M. SELIGMAN, HELPLESSNESS (1975) (depression linked to belief that one cannot control impor-

participating decisionmakers, but society as a whole, including Burt and his readers, feels better about decisions that are agonized over.

In-depth conversation may be psychologically important, then, even if the answer to the ultimate issues, or the ultimate issues themselves, are unknowable. Full consideration of the issues should be even more important to the extent that some of the crucial questions involved are either answerable or knowable. Surely some are: Has Mr. G been informed that psychiatrists believe he may wish to end his life not because of a "rational" desire to terminate pain but because of an "irrational" guilt over his father's death? If he is so informed, would that information influence his decision? Does Mrs. Lake's inability to recall her address mean that she cannot find her way home or does it mean only that she has difficulty remembering digits?[60] Could Karen Quinlan breathe without the respirator? Could that question be answered by removing her from the respirator for a few minutes?[61] Is Mr. Saikewicz so severely retarded that his acquiescence in chemotherapy could not have been obtained? Could his cooperation be obtained through heavy sedation and intensive staff efforts to calm him?[62]

Burt raises many of those important questions and is right to be concerned that they were not asked or answered before or during the pre-treatment adjudications in *Lake*, *Quinlan*, and *Saikewicz*. It is his thesis that the stress and destructive consequences unleashed by distressed patients and fueled by pre-treatment legal intervention makes it likely that these important matters will go unexplored in commitment hearings and actions for declaratory relief. He presumably believes that those same issues *would* be addressed in a legal system that generates communication by emphasizing uncertainty and the possibility of post-treatment legal redress.

Yet, the context in which conversation would occur under Burt's model possesses many, though not all, of the attributes that students of group dynamics suggest will produce "groupthink"[63] or

tant elements of one's life); Geer, Davison, & Gatchel, *Reduction of Stress in Humans Through Nonveridical Perceived Control of Aversive Stimulation*, 16 J. PERSONALITY & SOC. PSYCH. 731 (1970) (illusion of control over experimental shocks make shocks less stress-inducing).

60. *See* p. 26.

61. In fact, before the court hearing, she had been so removed uneventfully, but that fact was never forcefully argued. P. 154.

62. P. 157.

63. *See* I. JANIS & L. MANN, DECISION MAKING: A PSYCHOLOGICAL ANALYSIS OF CONFLICT, CHOICE, AND COMMITMENT 129-33, 395-400 (1977); Janis, *What Group Dynamics Can Contribute to the Study of Policy Decisions*, in POLICY STUDIES AND THE SOCIAL SCIENCES 125 (S. Nagel ed. 1975); Siegel & Zajonc, *Group Risk Taking in Professional Decisions*, 30 SOCIOMETRY 339 (1967).

other types of faulty group decisionmaking.[64] Groupthink is a concurrence-seeking tendency that impairs careful decisionmaking in cohesive groups.[65] Even in ad hoc groups with great internal conflict, members generally try to "smooth conflicts over rather than resolve them,"[66] a practice that often leads to "wildly inaccurate solutions."[67]

Group decisionmaking, which in most circumstances is superior to individual decisionmaking,[68] plummets in quality in the face of authoritarian leadership, crisis situations, stress, moral dilemmas, member dependence on group approval for maintaining a self-image of decency, and relative insulation of group members until after a final decision.[69] Consider those characteristics in relation to Burt's conversational model for a *Quinlan*-type situation: Doctors clothed in the authoritative and authoritarian mantle of medicine meet with family members at a time of crisis and stress to ponder the moral dilemma of continuing life-support systems, with each member seeking group concurrence in and support for a decision the decency of which will later be judged.

Students of group dynamics, and of decisionmaking generally, have devised methods to minimize many of the obstacles that impede accuracy in reaching decisions.[70] They have suggested, for example, that each member should be a "critical evaluator,"[71] and should encourage the group "to give high priority to airing objections and doubts."[72] Outside experts should be called in and "should be encouraged to challenge the views of the core members."[73] One or more members should be given the role of "devil's advocate," with an "unambiguous assignment to present his arguments as clearly and convincingly as he can, like a good lawyer, challenging the testimony of those advocating the majority position."[74] Finally, after reaching a "preliminary consensus," the group should hold a "second chance" meeting at which "every member is expected to express as vividly as he can all his residual

64. *See* Hall, *Decisions, Decisions, Decisions*, PSYCH. TODAY, Nov. 1971, at 51, 52-53.
65. *See* note 63 *supra* (citing sources).
66. Hall, *supra* note 64, at 52.
67. *Id.* at 53.
68. R. NISBETT & L. ROSS, *supra* note 49, at 267.
69. Janis, *supra* note 63, at 128-29.
70. R. NISBETT & L. ROSS, *supra* note 49, at 282-85; Hall, *supra* note 64, at 54, 86; Janis, *supra* note 63, at 130-31.
71. Janis, *supra* note 63, at 130.
72. *Id.*
73. *Id.* at 131.
74. *Id.*

doubts and to rethink the entire issue before making a definitive choice."[75]

If that advice is correct, which is not yet certain,[76] curtailing improper group decisionmaking procedures might be difficult under the relatively unstructured framework likely to emerge from Burt's after-the-fact legal framework. However stressful and destructive its context, a pre-treatment adjudicatory framework, with a judge focusing discussion and monitoring conflict, may be far more suitable for accommodating the above-suggested legally familiar devices for minimizing error in multi-party decisionmaking.

These seem to me to constitute some of the major conceptual and empirical issues that must be confronted in structuring a role for the law in medical and psychiatric decisionmaking. *Taking Care of Strangers*, a provocative book full of creative scholarship, will make a major contribution to this important field by forcing a reexamination of the role of law in it.[77] If these matters are at all knowable, the intellectual excitement and agony spawned by Burt's work should improve our decisions regarding appropriate legal structures for doctor-patient relationships.

75. *Id.*

76. The proposed remedies have yet to be empirically validated. I. JANIS & L. MANN, *supra* note 63, at 400. Somewhat similar procedures have, however, withstood experimental scrutiny. Hall, *supra* note 64, at 88.

77. As this review goes to press, some five months after grappling with the book, I am struck by how much my own thinking has been influenced by Burt's central point regarding the stress and discomfort generated by patients in pain. Whether or not his solution is accepted, his exposition of the problem may itself constitute a major and lasting contribution to legal scholarship in this field.