

Rethinking Expert Testimony in Education Rights Litigation

Kevin G. Welner

University of Colorado at Boulder

Haggai Kupermintz

University of Haifa, Israel

Courts often rely on the testimony of experts to understand arguments and implications in education rights litigation. But expert testimony, and statistical testimony in particular, can offer a false sense of security for the unwary. This article uses expert testimony offered in two recent desegregation cases to consider whether sufficient protections are presently in place to protect judges, who are usually statistical novices, from being confused or misled by experts. These case studies illustrate how, without the use of additional protections, courts can be misled. Following this examination, we offer suggestions intended to improve judges' comprehension of expert testimony. At its most general level, this article addresses the role of researchers in presenting important educational issues in ways that speak clearly to policy-makers.

Keywords: *ability grouping, expert witnesses, law, policy, tracking*

THE TESTIMONY of experts can be, and has been, indispensable in helping courts understand arguments and implications in education rights litigation. The social science evidence presented in *Brown v. Board of Education* (1954) was a watershed; over the succeeding five decades, courts have repeatedly looked to social science and statistical experts for assistance in determining whether past wrongs have been remedied and whether new wrongs have been committed. As the Supreme Court stated, in the context of employment discrimination, "Since the passage of the Civil Rights Act of 1964, the courts have frequently relied upon statistical evidence to prove a violation. . . . In many cases the only available avenue of proof is the use of racial statistics to uncover clandestine and covert discrimination by the employer or union involved" (*International*

Brotherhood of Teamsters v. United States, 1977, p. 339–40, fn. 20).

Most recently, the *amicus curie* brief filed by the American Education Research Association in the Michigan affirmative action case of *Grutter v. Bollinger* (2003), was cited approvingly by the court for the proposition that "numerous studies show that student body diversity promotes learning outcomes, and 'better prepares students for an increasingly diverse workforce and society, and better prepares them as professionals' " (p. 333). But expert testimony, and statistical testimony in particular, can offer a false sense of security for the unwary. As the Supreme Court has warned, "Expert evidence can be both powerful and quite misleading because of the difficulty in evaluating it" (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1993, p. 595, internal citation omitted).

Dr. Welner's research for this article was funded, in part, through a postdoctoral fellowship from the Spencer Foundation and the National Academy of Education. We would also like to thank Ronald Allen, Damian Betebenner, Carol Burris, David Faigman, Gilbert FitzGerald, Gene Glass, David Kaye, Mike Rebell and Alison Vredenburg for their insights. The opinions and ideas expressed herein are solely those of the authors.

The Court has also cautioned that expert testimony has a heightened potential to sway a jury “because of its aura of special reliability and trustworthiness” (*United States v. Amaral*, 1973, p. 1152). At about the same time, Hank Levin and Willis Hawley issued a similar warning in the context of education, pointing to a “real possibility . . . that substantial misinterpretation and overstatement will accompany the increasing use of social science evidence in the courts” (Levin & Hawley, 1975, p. 3).

This article explores the use of expert testimony in two recent desegregation cases. The authors were among the plaintiffs’ expert consultants in these cases. In conjunction with unitary status hearings in those cases, experts offered analyses to support or refute the claim that the district no longer engaged in discrimination against minority students. (Courts hold unitary status hearings to determine whether the school district is no longer operating a dual system, one for White students and one for Latino or African American students.) If the dual system has been replaced by a unitary one, and the school district is said to have reached “unitary status.” In particular, the experts addressed the issue of potential resegregation of students through the use of ability grouping, or “tracking” (see Welner & Oakes, 1996). We use the statistical testimony offered in these recent desegregation cases to illustrate how, without the use of additional protections, courts can be misled. In the context of these case studies, we explore the protections and mechanisms currently in place to assist judges and juries in their understanding of complex issues. Following this examination, we offer options for improving judges’ ability to competently evaluate expert testimony.

At its most general level, however, this article addresses the role of researchers in presenting important educational issues in ways that speak clearly to policy-makers. Issues of comprehension pervade the policy-making arena. In one recent example, newspapers and politicians enthusiastically grabbed onto a controversial study concerning the impact of Florida’s voucher system (Greene, 2001). The study was prominently hailed as providing empirical support for federal voucher legislation (see Henry, 2001; Schemo, 2001), even though the findings it reported could be attributed to regression toward the mean plus drilling for a specific component of the test (Kupermintz, 2001; see also Camilli & Bulkley, 2001). Such

examples highlight the broader need for more training and preparation of experts for communication in each type of policy-making arena.

Evidence from the Desegregation Cases

In two recent desegregation cases,¹ experts were called upon to provide evidence pertaining to the question of bias in student access to advanced courses. The school districts in these desegregation cases are located in Rockford, Illinois and Wilmington, Delaware. These are both mid-sized cities whose populations have remained fairly stable over the past few decades, even under court-ordered desegregation. Unlike bigger cities, where White and middle class flight have made integration so difficult, these communities maintained racially-mixed schools with sizable middle class participation.

The desegregation cases in Rockford and Wilmington are described in detail in Welner (2001). In a nutshell, both school districts had well-documented histories of intentional racial segregation. They both were placed under court desegregation orders. In both cases, the desegregation plaintiffs called the court’s attention to an indisputable pattern: significantly fewer racial minority students were enrolled in higher-tracked courses, as compared to their majority counterparts. Pointing to this disparity, the plaintiffs alleged that the districts used tracking as a form of so-called ‘second-generation’ discrimination in response to court-ordered between-school desegregation. That is, they contended that tracking subverted between-school integration by separating students within the school site. Experts offered statistical analyses to address the question of whether this enrollment pattern indicated racially discriminatory placement practices or whether it simply reflected differential levels of academic preparedness of minority and majority students. The latter explanation of the observed enrollment patterns might constitute, as a legal matter, an inevitable consequence of a rational and defensible educational policy.

Each party’s experts considered three variables: previous year’s standardized test scores on the SAT9 or ITBS (depending on the district), whether the student was enrolled in an advanced course, and student race. The latter variable was dichotomously stated as either “minority” or “majority.” The courts used these terms for added flexibility, since “minority” sometimes included African Americans, sometimes Latinos, and sometimes both, while “majority” sometimes

included Whites alone but sometimes also included Asian Americans.

Analyses were conducted for courses in English, math, science, reading, and social science. Experts for the defendants (the school districts) and plaintiffs presented the court with sharply different analytic approaches to the problem, leading to conflicting conclusions about the extent to which enrollment patterns reflected discriminatory placement practices. These cases were all tried by judges, without juries. Thus, it fell upon the judges to weigh the evidence and apply the expert analyses to decide whether course enrollment was tainted by racial bias.

Approach A

The expert for the defendants was a political scientist who has published extensively on the issue of desegregation. The courts qualified her to testify about the course enrollment process in each district and how that process affected racial segregation. She offered the following analysis. At the most basic level, she compared two proportions: (a) the proportion of students enrolled in high school advanced classes who are minority students, and (b) the proportion of all high-scoring students (students scoring above, e.g., the 75th national percentile on the SAT9 or ITBS in the year prior to placement) who are minority students. The rationale behind this comparison was predicated on the assertion that enrollment in higher-tracked classes is rigidly linked to academic merit and therefore that high-scoring students are expected to enroll in those advanced courses. Thus, a comparison of the racial composition of the high-scoring student group to the racial composition of students who actually enrolled in advanced courses should reveal whether the rate of minority enrollment in advanced courses matches the proportion that would be predicted on the basis of academic merit (as evidenced by high test scores) alone. For example, in a hypothetical school with 100 students, 50 minority and 50 majority, assume that 25 students score above the 75th national percentile: 5 minority and 20 majority. Assume that the high-track class includes 25 students: 7 minority and 18 majority. The expert would compare the second fraction (7/25) to the first (5/25) and conclude that minority students are over-represented in the high-track class.

In the actual cases, the general finding of the defendants' expert was, in fact, that proportion (a) was greater than proportion (b). Formally, in

terms of conditional probabilities, the defendants' expert analysis purported to interpret the fact that $Pr\{\text{minority} \mid \text{advanced course}\} > Pr\{\text{minority} \mid \text{high achievement}\}$ as evidence that the district was in fact implementing an affirmative action policy, since more minority students were actually enrolled in advanced courses than was justified by their test scores. Results from the Rockford case are presented in Figure 1, which reproduces a figure presented in the report of the defendant's expert.

Approach B

The expert for the plaintiffs in these cases was an educational researcher who has published extensively on the topic of tracking. The courts qualified her to testify about the same issues as the defendant's expert: the course enrollment process in each district and how that process affected racial segregation. She performed an analysis using ANCOVA (Analysis of Covariance) methodology, whereby she examined the likelihood of placement in high school advanced courses as a function of race, after controlling for achievement as measured by test scores in the year prior to placement. In the discrete version of the analysis, the distribution of test scores was divided into 10 sub-samples of roughly comparable achievement, each containing 10% of the data. The experts found that within each sub-sample minority students were less likely to be enrolled in advanced courses. The continuous version of this same type of analysis uses logistic regression to conclude that $Pr\{\text{advanced course} \mid \text{minority}\} < Pr\{\text{advanced course} \mid \text{majority}\}$, controlling for achievement level. That is, student race accounted for segregatory patterns of enrollment in advanced courses, over and above that which could be predicted on the basis of test scores alone. These results were interpreted to support the conclusion that advanced course enrollment was indeed biased against minority students. Results from the Rockford case, presented as discrete sub-samples, are presented in Table 1. Results based on a logistic regression model are presented in Figure 2.

Clearly, the two sets of analyses and findings pointed the experts to opposing conclusions. The courts were accordingly faced with the task of determining which of the two conclusions was more appropriate given the probative value of the evidence, analyses, and interpretations presented. In doing so, the judges first had to recognize that the two analyses addressed very differ-

**“Ratio of Honors Enrollment Racial % to High Scorers*
Racial % for High School Students, Rockford, Fall 1999”**

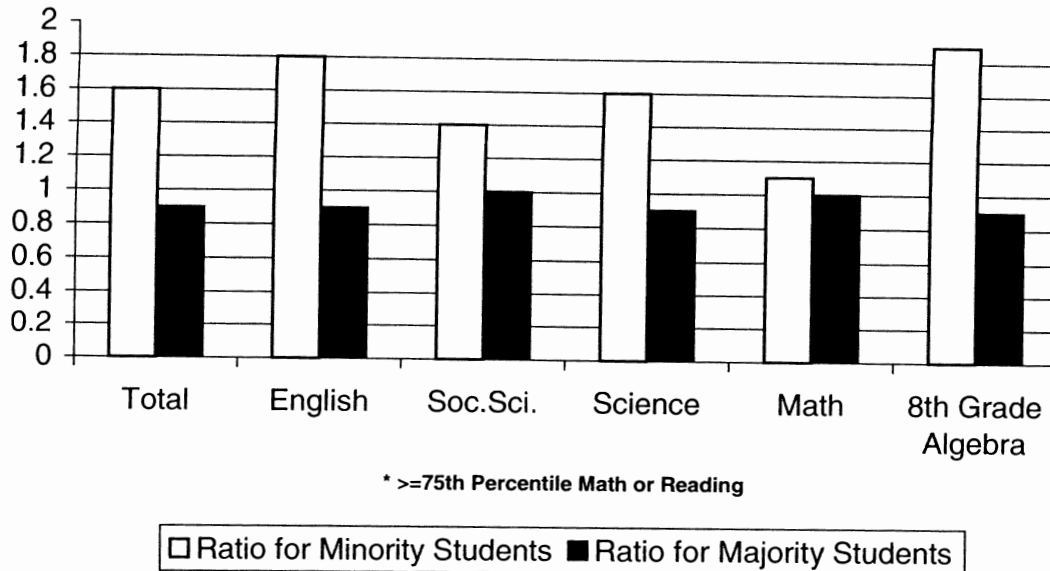


FIGURE 1. *Defendants' Expert's Figure 22 in her Rockford Report.*

ent empirical questions and then had to evaluate each question's appropriateness for addressing the issue before the court: Did the school district's course placement practices discriminate against

minority students? Approach A asked, *Were minority students represented in advanced courses in accordance with their proportion among high-scoring students?* In contrast, Approach B asked,

TABLE 1
Rockford Placement of Majority and Minority High School Students at Each "Slice" of Math/Reading Achievement in Regular and Advanced Classes—1999–2000

Math/Reading Achievement	Majority Students	Minority Students
First (Lowest)	633* Percent Advanced: 3%**	1212 Percent Advanced: 2%
Second	941 Percent Advanced: 5%	1207 Percent Advanced: 4%
Third	1166 Percent Advanced: 7%	1103 Percent Advanced: 7%
Fourth	1492 Percent Advanced: 15%	1109 Percent Advanced: 13%
Fifth	1367 Percent Advanced: 20%	667 Percent Advanced: 15%
Sixth	1819 Percent Advanced: 31%	650 Percent Advanced: 24%
Seventh	1788 Percent Advanced: 46%	446 Percent Advanced: 37%
Eighth	2346 Percent Advanced: 61%	377 Percent Advanced: 59%
Ninth	2271 Percent Advanced: 74%	227 Percent Advanced: 70%
Tenth (Highest)	2124 Percent Advanced: 86%	9 Percent Advanced: 81%

Notes. *Meaning that 633 majority students scored in the lowest slice. **Meaning that 3% of these 633 majority students (that is, 19 students) were enrolled in the advance track.

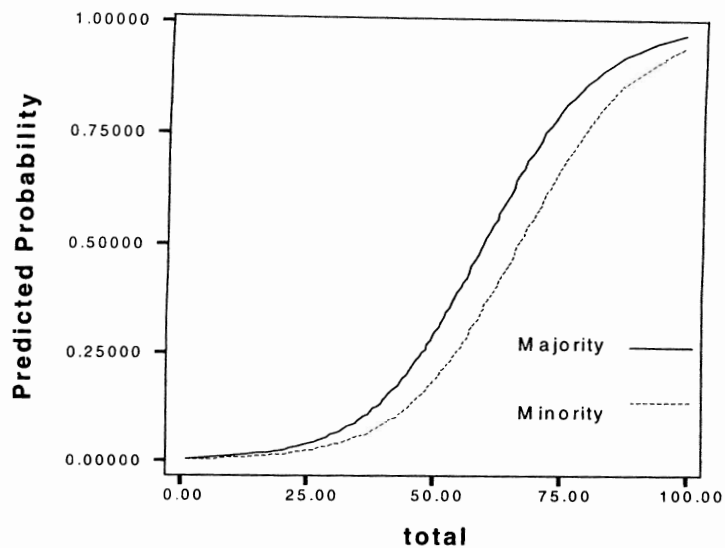


FIGURE 2. Rockford placement of majority and minority high school students as a function of math/reading achievement in regular and advanced classes—1999–2000.

For students with similar levels of achievement, was the rate of enrollment in advanced courses the same for minority and majority students?

In evaluating the probative value of the analyses, the courts were further called upon to consider the implicit assumptions underlying the two approaches. Both analyses used standardized test scores as the only criterion against which racial variations in course placements were examined. These test scores are subject to measurement error, which may have accounted for some of the observed differences. Also, the actual placement process was far more complicated and included a variety of criteria, among which test scores were only one component. Neither approach addressed this key limitation; the use of test score benchmarks as the eligibility criterion for enrollment in advanced courses therefore results in conclusions that can be misleading. This limitation has to be acknowledged and evaluated in order for a court to understand the complex context of course assignment practices and the usefulness of the experts' analyses in disentangling the direct effects of ethnicity from other factors such as parental involvement, teacher-student relations, or accumulated social capital. Race is strongly associated with test scores as well as these other factors (Mickelson, 2001; Welner, 2001).

Another issue facing the courts concerned the practical significance of the analyses. For instance, although analyses offered by Approach B indicated differential rates of placement into advanced courses, the expert provided little guidance for

the courts to evaluate the severity of the problem. No criteria were suggested that would help the court evaluate the 'effect size' of the nominal disparities (see Oakes, 1995, 2000). Effect size issues are important as a policy matter, but for the analyses discussed here they are not important in terms of statistical significance. The analyses used population data, rather than sampling, and the experts were not interested in generalizing the findings to a larger population.

Perhaps the most difficult statistical issue presented in these analyses concerns a unique statistical paradox inherent in Approach A—similar to, but more complex than, the well-known Simpson's Paradox (Simpson, 1951; see also Appleton, French, & Vanderpump, 1996; Bickel, Hammel, & O'Connell, 1975; Mantel, 1982; and Wainer, 1986).² The Approach A paradox results because the majority and minority students' test score distributions in these districts reflected the national trend of divergence (of approximately one standard deviation), and, for each racial/ethnic group, the subpopulation of students placed in the advanced classes was not necessarily drawn from the subpopulation of high-scoring students. The approach ignores the extent to which the high-scoring students in each group were actually placed in advanced classes, or the extent to which low-scoring students were placed in regular classes. Even though the percentages of 'high scorers' and the percentages of students placed in advanced classes appear to advantage minority students over majority students in the Approach A

analysis (indicating no discrimination), the *rate* at which high-scoring minority students were overlooked for advanced class placement is greater than for majority students (evidence of discriminatory placement). Further, the rate at which low-scoring minority students were placed in advanced classes is also lower than for majority students.

Without comparable groups, large numbers of majority students can be added to the advanced classes and be 'balanced' by adding just a few minority students. Because of the distinct score distributions of the two racial populations, this resulting statistical equality exists side-by-side with a racially discriminatory pattern of class placement across each slice (Table 1). The Approach A comparison can paradoxically show no discrimination or even what appears to be an affirmative action effect (Figure 1). The seemingly equitable result is that overall enrollment in higher tracked classes can show a racial balance stronger than one would expect if the classes simply enrolled the top-scoring 25%.

Furthermore, in evaluating the probative value of the experts' analyses, the court was called upon to be aware of the fact that Approach A excluded most of the available data from the analysis (for example, approximately 70% of the data in the Wilmington and Rockford cases) and accordingly ignored crucial distributional differences between the compared groups. Returning to the simplified, 100-student example used earlier, any student not among the 25 placed in the high-track class and/or the 25 scoring above the 75th national percentile was omitted from the Approach A analysis. Perhaps more importantly, if the student qualified for one group but not the other, then she or he was considered in only one part of the comparison. In addition, the outcome of Approach A strongly depends on the specific cut-point (the 75th percentile) used to define "high achievement." Varying the cut point results in the inclusion of either more or less of the data and can cause the evidence for or against discrimination to change dramatically.

In sum, potential limitations of the analyses submitted to the courts included: (a) failure to consider the entire range of ability, (b) false determinism regarding the relationship of test scores to course placements, and (c) the absence of information on the magnitude of the reported effects in the analyses. The adversarial process in the two cases certainly produced strong critiques of both approaches, including many of the above-discussed issues. But neither responsive expert reports nor

cross-examination of the experts were sufficient to assist the judges with their understanding of the analyses. In the next section of this article, we offer the case study of Wilmington as an example of the difficulties faced by courts trying to make sense of complicated statistical analyses and arguments.

A Chronology of Confusion

As noted earlier, each of the two approaches described above was offered and accepted as evidence in at least two desegregation cases. In one of these cases (Rockford), the court dismissed Approach A, accepting Approach B.³ In the other case (Wilmington), the court included in its factual findings conclusions based on both analyses, but found in favor of the defendant school districts (see *Coalition to Save our Children v. State Board of Education*, 1995, pp. 800–801).

The Wilmington decision produced a powerful demonstration of the difficulties faced by many courts and attorneys when tackling statistical evidence. The report from the defendants' expert witness included analyses conducted in accordance with Approach A as explained above. It presented a chart (reproduced here as Figure 3) and explained this figure as follows: "This chart shows the percentage of high school students by race scoring at or above the 75th percentile in math or reading enrolled in one or more honors or advanced placement (AP) classes" (Rossell, 1994, p. 29). The chart alluded to (see Figure 3) displays the standard comparisons of Approach A, this time cast in terms of percentages: $[Pr\{\text{minority} | \text{advanced course}\} / Pr\{\text{minority} | \text{high achievement}\}] * 100$ (with a similar calculation for majority students). Again using the 100-student school example, the calculation might proceed as follows: $[(7/25)/(5/25)] * 100 = 140\%$.

However, the expert report's description of the figure lends itself to a different interpretation, suggesting that the percentages exhibited are those of minority and majority students in advanced classes, out of the group of high-scoring students: "% of High School Students Scoring at or Above the 75th Percentile on Reading or Math, Spring 1993 Enrolled in Honors/AP Courses, Spring 1994." That is, among the subgroup of students who are high scoring and are also among the students placed in the advanced class, what percentages are of each race? On careful consideration, this interpretation is belied by the fact that the percentages in the chart exceeded 100 in many cases, and the chart's footnote explained that in some

“% of High School Students Scoring at or Above the 75th Percentile on Reading or Math, Spring 1993 Enrolled in Honors/AP Courses, Spring 1994^a”

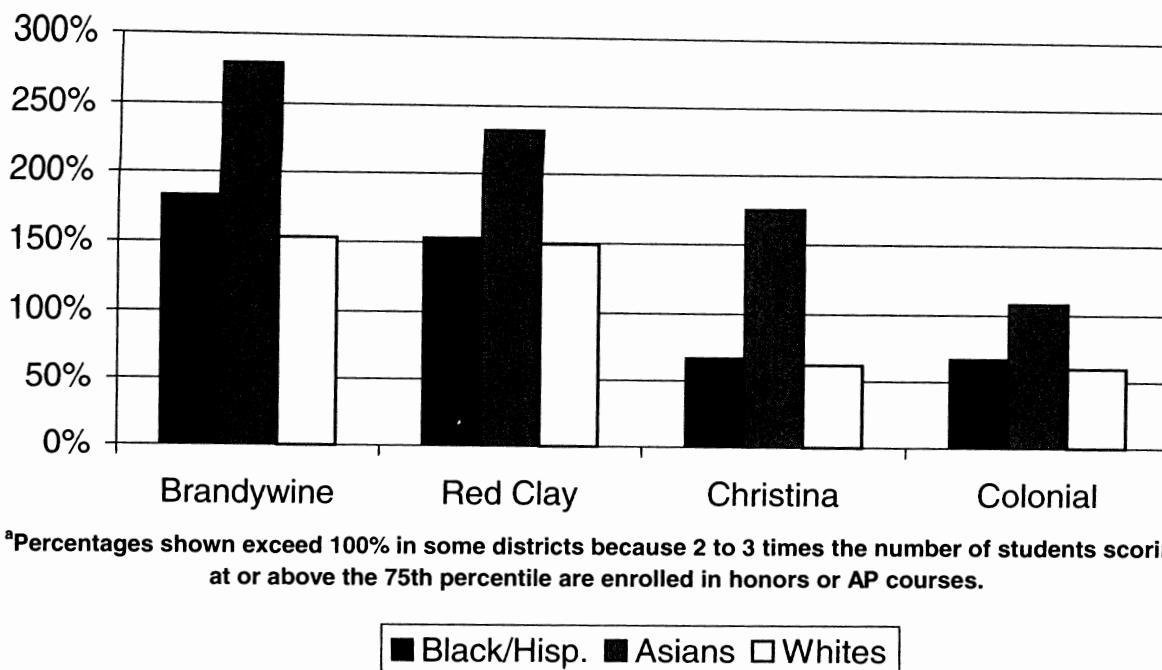


FIGURE 3. *Defendants' Expert's Figure 13 in her Wilmington Report.*

cases: “. . . 2 or 3 times the number of students scoring at or above the 75th percentile are enrolled in honors or AP courses,” thus indicating a ‘classic’ Approach A calculation.

Yet the ambiguity of the expert witness’s explanation engendered, for the trial court, manifest confusion as to the nature of the evidence. The following factual finding (called “Finding 48” in the written opinion) emerged: “the percentage of minorities enrolled in honors and AP classes who scored over the 75th percentile in reading or math in the spring of 1993 is slightly greater than that of Whites . . . [citation to expert report]” (*Coalition to Save our Children v. State Board of Education*, 1995, p. 801). The court interpreted the expert’s figure and explanation as referring to the proportion of high-scoring minority and majority students out of those enrolled in honors or AP classes, or *Pr*{high achievement | advanced course}, a quantity never calculated by Approach A or presented to the court as evidence.

This erroneous factual finding then served as a basis for an appellate-level debate that is interesting but is simply not grounded in any valid evidence. Finding 48—which was already a misinterpretation of the defendants’ expert’s statement of her own paradoxical analysis—was now turned on its head; it became evidence

for rather than *against* potential racial discrimination in course placements. Both the majority and dissent read the finding to mean that among students placed in the high-track classes, the minority students had higher test scores, thus implying that the White students were admitted even though less qualified. Ironically, this is similar to the opinion offered by the plaintiff’s expert, based on Approach B. In upholding the judgment in favor of the defendant school districts, the appellate majority minimized the importance of this freshly interpreted finding that now pointed in the opposite direction. The majority therefore reasoned: “Although this finding could, as urged by Appellant, give rise to an inference that Blacks must perform at a higher level than Whites in order to be placed in honors and AP classes, that is not the sole inference that could be drawn from so limited, and thus malleable, a sample. Indeed, on the basis of finding 48 alone we may just as reasonably infer something quite different: that the school districts’ good faith efforts to desegregate have paid off in terms of the improved testing performance of Black students” (*Coalition to Save our Children v. State Board of Education*, 1996, p. 764). The dissenting appellate opinion countered: “The Majority’s inference, even assum-

ing that it is one an appellate court could draw, is unconvincing at best: the issue is not whether some Black students perform well, but rather whether Black students must perform better than Whites to be placed in honors and AP classes, which the district court's finding clearly suggests" (*Coalition to Save our Children v. State Board of Education*, 1996, p. 783). As this chain of confusions and misinterpretations demonstrates, the probative value of the statistical analysis presented to the court completely vanished and led both the trial court and the appellate panel to consider irrelevant arguments, pro and con, based on misinterpreted analyses and evidence.

The Admission of Expert Testimony

Expert testimony occupies a special niche within the larger body of trial testimony. Experts are called on to provide specialized knowledge concerning issues and concepts about which the trier of fact must decide yet presumably lacks the requisite knowledge to competently evaluate. The trier of fact may be a jury or, in a so-called "bench trial," a judge. The trier of fact is often presented with conflicting testimony and is responsible for making determinations concerning the truth of what happened.

Experts testify to matters thought to be beyond the unaided power of laypeople to resolve. Because experts are not 'percipient witnesses'—meaning that they are not testifying about their first-hand observations and perceptions of matters in dispute—their testimony is often akin to garden-variety hearsay.⁴ Nonetheless, their testimony is welcomed for the purpose of explaining evidence that might otherwise be difficult, if not impossible, for the trier of fact to understand.

As discussed below, most of the concern about the potential of expert witness testimony to confuse or mislead has focused on juries, not judges. Yet the experiences of the judges in the Rockford and Wilmington cases suggest that this concern should extend to those cases where judges serve as triers of fact, as well as to the role of judges in screening evidence for admissibility in jury trials. For actions in equity, such as the desegregation lawsuits discussed in this article, there exists no general federal right to a jury trial. In these cases, the trier of fact is almost always a judge.

In the discussion that follows, we analyze the admissibility of expert witness testimony keeping in mind that the potential for confusion exists whether the trier of fact is a judge or a jury. In both types of cases, we consider the issue of evidentiary admissibility in the context of our apprehensions about judicial comprehension of expert testimony. If a jury hears the case the judge must competently evaluate the testimony in order to screen out confusing or misleading testimony. If a judge hears the case, without a jury, the judge must still competently evaluate the testimony. In these latter cases, the decision of whether or not to admit the testimony is largely irrelevant—if the judge is capable of soundly evaluating meritless testimony, she will give it no credence whether admitted or not. Judicial comprehension is crucial in each circumstance, so we will return to this comprehension issue throughout the discussion of admissibility.

In order to gain the benefit of expert testimony while reducing its potential harm, courts have applied various evidentiary rules, all designed to ensure that the trier of fact relies on trustworthy and appropriate evidence. However, the focus has overwhelmingly been on trials where a jury serves as the trier of fact; the rules rely upon the judge's comprehension of expert testimony in order to successfully screen out confusing or misleading testimony before it reaches that jury. The following overview of those rules demonstrates this reliance and highlights the importance of increasing judges' understanding of expert witness testimony.

Prior to 1975, when Congress enacted the Federal Rules of Evidence, the party offering testimony had the burden of demonstrating that the expert's testimony was based on a generally accepted theory or technique in the scientific community to which it related (*Frye v. United States*, 1923). This *Frye* test was criticized because newer scientific theories are generally less accepted but not necessarily less reliable. Moreover, some critics questioned the ease and value of "nose counting" among the scientific community (see Brown, 1999, at 779).

The new Federal Rules did not explicitly address the old *Frye* rule, but they did include the following provision concerning the testimony of experts: "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in

issue, a witness qualified as an expert by knowledge, skill, experience, training or education, may testify thereto in the form of an opinion or otherwise" (Rule 702). For the next eighteen years, from 1975 through 1993, courts varied in their application of *Frye*, Rule 702, or a combination of the two. But in its 1993 *Daubert* decision, the Supreme Court expressly held that the Federal Rules of Evidence superseded the *Frye* rule, and (as discussed below) the Court offered some guidelines for applying Rule 702. (Since state courts are not governed by the Federal Rules, *Frye* survives in many state courts. However, the emphasis of the analyses in this article is on federal courts.)

The *Daubert* case focused on the admissibility of 'scientific' expert testimony. The case involved birth defects allegedly caused by Bendectin, a drug once prescribed to treat morning sickness during pregnancy. The Supreme Court ruled that scientific expert testimony is admissible only if it is both relevant and *reliable* (citing Rule 702). It offered four specific factors as "general observations" for assessing scientific reliability (described on pages 593–594 of the Court's opinion).

1. Whether the theory or technique can be (and has been) tested. A method that is testable and therefore subject to falsifiability points toward reliability.
2. Whether the theory or technique has been subjected to peer review and publication.
3. The known or the potential rate of error for any tests or techniques and the existence of standards or controls for the technique's operation.
4. Whether the theory or technique has been generally accepted.

These *Daubert* guidelines require judges to engage in a more active role as gatekeeper for expert witness scientific testimony.

The Court clarified *Daubert* in two later cases, increasing its scope and flexibility (*General Electric Co. v. Joiner*, 1997; *Kuhmo Tire Co. v. Carmichael*, 1999). In particular, the Court clarified two matters: (a) *Daubert's* general holding—setting forth the trial judge's gatekeeping obligation—applies not only to testimony based on "scientific" knowledge, but also to testimony based on "technical" and "other specialized" knowledge (*Kuhmo*, p. 151); and (b) the test of reliability is flexible, with the four factors listed in *Daubert* not necessarily being useful or

applicable to all experts or in every case (*Kuhmo*, p. 153). The Court also instructed lower courts not to admit or accept "opinion evidence which is connected to [the] data only by the *ipse dixit* [unproven assertion] of the expert" (*Joiner*, 1997, p. 146). A consequence of these two cases has been the effective elevation of the judge's gatekeeping role over the effort to liberalize the admissibility of scientific evidence (see Imwinkelried, 2000). In theory, this should assist the jury by filtering out less reliable testimony.

Technically, Rule 702 also requires the judge in a bench trial to play the role of gatekeeper toward evidence that only she will consider. But, as noted earlier, this issue of exclusion of expert testimony has importance in jury trials but is largely irrelevant in bench trials.

In 2000, Federal Rule 702 was amended in response to these cases. The amendment consists of the addition of three conditions, focusing on the principles of sufficiency and reliability, which must exist in order for an expert to be allowed to testify. As amended, Rule 702 reads:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data; (2) the testimony is the product of reliable principles and methods; and (3) the witness has applied the principles and methods reliably to the facts of the case.

Note that, in all three cases discussed above, the Supreme Court was not concerned about judges being confused or misled by expert testimony. The concern was the possibility of a *jury* being misled. This system seems well designed if the expert testimony falls in the zone between the competence of the jury and the judge (see Figure 4). However, the Supreme Court's (and Federal Rules') reliance on judges to screen out unreliable or misleading expert testimony failed to address two circumstances: (a) when the judge, in trial with a jury, is not competent to evaluate the testimony; and (b) when the judge, in a bench trial, lacks this competence. In the former situation, the jury will be presented with testimony that, perhaps, should have been screened out. In the latter situation, the role of the judge as gatekeeper of admissibility becomes moot, since a judge's evaluative incapacity

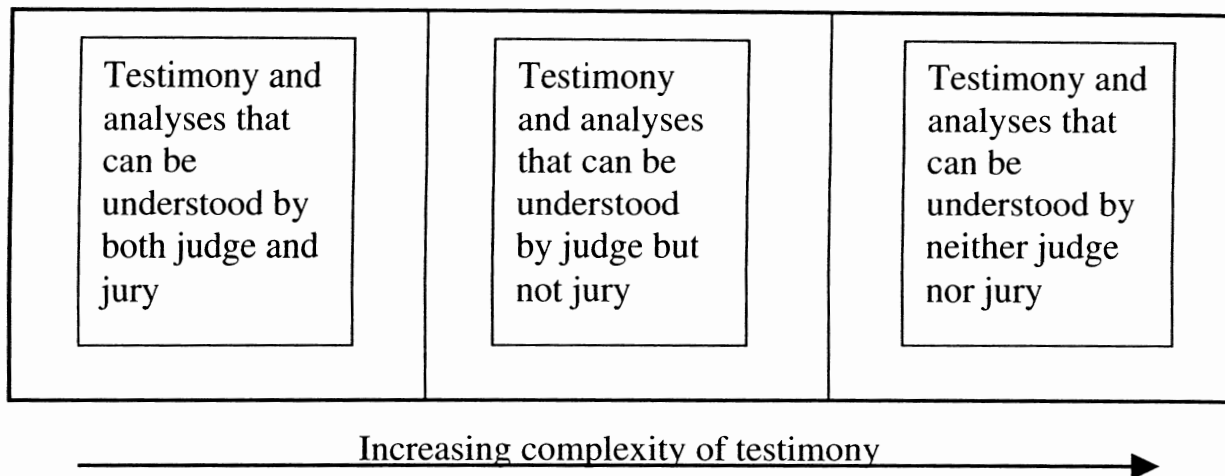


FIGURE 4. Range of understandability of testimony.

ity becomes an equivalent liability whether deciding on admissibility or deciding on the merits of the expert’s analyses and conclusions.

The Dilemma Posed by Confusing or Misleading Expert Testimony

The elevated gatekeeping role has given judges the authority, but not necessarily the ability, to screen out expert testimony that is “misleading” or “confus[ing]” (Federal Rule 403). This issue was tangentially addressed by the *Daubert* Court, which admonished lower courts to be mindful of Federal Rule 403 as well as Rule 702. Rule 403 provides that relevant evidence may be excluded if its probative value is substantially outweighed by “confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative

evidence.” Relevant evidence means evidence having a tendency to make the existence of any consequential fact (i.e., of consequence to the determination of the action) more probable or less probable than it would be without the evidence.

Under Rule 403, the *Daubert* Court advised, trial courts should exercise “more control over experts than over lay witnesses” (p. 595). Accordingly, satisfying the *Daubert* reliability test does not necessarily mean that expert testimony will satisfy Rule 403 (see, e.g., *United States v. Hicks*, 1996, p. 847; *Guillory v. Domtar Indus., Inc.*, 1996, p. 1331; see also Brown, 1999).

Rule 403, however, provides a rather limited remedy (see Table 2, offering a comparison of the two rules). It permits the exclusion of scientific evidence based upon the potential for confusion of the trier of fact *only if* it’s a jury trial and if there

TABLE 2
Comparison of Federal Rules 403 and 702

Rule 403	Rule 702
Applies only to jury trials.	Applies to all triers of fact (judge or jury).
Provides that relevant evidence may be excluded if its probative value is substantially outweighed by “confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.”	Authorizes the admission of expert witness testimony if “scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or determine a fact in issue.”
Permits the exclusion of scientific evidence based upon the potential for confusion of the trier of fact <i>only if</i> there is “something particularly confusing about the scientific evidence at issue—something other than the general complexity of scientific evidence.”	Amendment to the Rule in 2000 notes three specific criteria to be considered by the judge in making the admissibility decision: (1) whether the testimony is based upon sufficient facts or data, (2) whether the testimony is the product of reliable principles and methods, and (3) whether the witness has applied the principles and methods reliably to the facts of the case.

is “something particularly confusing about the scientific evidence at issue—something other than the general complexity of scientific evidence” (*In re Paoli R.R. Yard PCB Litig.*, 1994, p. 747). Moreover, courts are averse to exclude relevant expert testimony, relying instead on the adversary process of cross-examination and refutation to expose weaknesses (*United States v. Baller*, 1975, p. 466). Such reliance may be a mistake; the admission of misleading expert testimony cannot necessarily be remedied through the adversary process:

Though there is some disagreement, most commentators believe ostensibly scientific testimony may sway a jury even when as science it is palpably wrong. Science can be greatly distorted by the pressures of litigation, but once admitted into evidence, it has an imprimatur of legitimacy and validity, and cross-examination often will not expose its flaws. Veracity, memory, motivation, prejudices, and biases—the weaknesses that cross-examination is best at ferreting out—are not very relevant to attacking an expert’s reasoning, and even lawyers who understand the flaws in an expert’s testimony may have difficulty explaining them to a jury. Experienced expert witnesses know all too well how to hide the ball. Many lawyers hesitate even to try cross-examining experts, because if they fail they only make things worse (Black, Ayala, & Saffran-Brinks, 1994, pp. 789–790; citations omitted).

Cross-examination also will not necessarily protect a *judge* from the dangers of misleading expert testimony, particularly in situations where the complexity of the testimony reduces the judge to weighing the *curricula vitae* of the experts in lieu of weighing the merits of their analyses.

The example provided by the above-discussed desegregation cases highlights a dilemma that is perhaps less confusing but no less perplexing than the statistical paradox discussed earlier. Given that a judge might be misled by certain complex testimony, how much can realistically be accomplished by Federal Rules 702 and 403 placing the onus on that same judge to somehow pre-empt such confusion by ferreting out (and excluding) that testimony before it reaches a jury? Upon remand of *Daubert* back to the Ninth Circuit Court of Appeals, the panel of judges expressed some dismay over the impossibility of the task:

[T]hough we are largely untrained in science and certainly no match for any of the witnesses

whose testimony we are reviewing, it is our responsibility to determine whether those experts’ proposed testimony amounts to “scientific knowledge,” constitutes “good science,” and was “derived by the scientific method.” . . . Our responsibility, then, unless we badly misread the Supreme Court’s opinion, is to resolve disputes among respected, well-credentialed scientists about matters squarely within their expertise . . . [W]e take a deep breath and proceed with this heady task (*Daubert v. Merrill Dow Pharmaceuticals, Inc.*, 1995, p. 1316).

At its root, this dilemma is engrained in the role the American judicial system has given to experts. As explained a century ago by Judge Learned Hand in a comment focusing on jury competence: “The whole object of the expert is to tell the jury, not facts . . . but general truths derived from his specialized experience. But how can the jury judge between two statements each founded upon an experience confessedly foreign in kind to their own? It is just because they are incompetent for such a task that the expert is necessary at all” (Hand, 1901, p. 54).

Here, it is important to distinguish between two possible judicial actions pursuant to the Federal Rules. The judge may (a) keep testimony from the jury under Rule 403 because the court knows that the testimony is simply too *confusing* for laypeople (perhaps including the judge) to understand; or (b) keep testimony from the jury under Rule 403 because it is so *misleading* that even the judge has been misled. The judge can clearly do (a). But one must strongly sympathize with a judge expected to do (b). If the testimony appears on its face to be understandable, reliable and probative, then the safeguards under Rule 403 add nothing to those under Rule 702. If, on the other hand, the judge recognizes the testimony as deceptive, then there is no need to look to Rule 403, since the judge will exclude it under Rule 702.

Of course, the intent of the expert is not our focus when we discuss “misleading” or “deceptive” testimony. Instead, these terms should be understood to mean *confusing testimony resulting in erroneous understandings and conclusions by the trier of fact*. Although an expert may feel pressure to come up with analyses and conclusions that support the client’s position, this will rarely drive the expert to intentionally mislead a court. We presume no wrongful intent on the part of any experts or testimony discussed in this article.

Judicial Options

In this concluding section, we examine possible approaches to providing judicial supports. For the benefit of those readers who serve as experts, we have not shied away from including some details concerning judicial procedures.

The desegregation cases discussed earlier provide an example of how this system can fail. Two federal district courts heard comparable testimony and reached opposite conclusions. Although this fact alone is not enough to indicate a system failure, what is particularly troubling is that *neither court demonstrated understanding in evaluating the analyses*. Moreover, while these were bench trials, a jury would likely have been equally—if not more—confused, and the judges would likely have been unable to adeptly evaluate the testimony for purposes of determining admissibility. Of course, the testimony offered in these cases was potentially understandable. Some judges will understand more than others. But we are inclined to sympathize with the judges in these cases; a tremendous amount was expected of them, with little preparation and few tools at their disposal.

Looking specifically at Approach A, the Rockford court rejected the testimony, but not because it identified any flaws in the statistical analysis.⁵ In the Wilmington case, the court adopted findings from both analyses—those based on Approach A and those based on Approach B—even though the two approaches could not be reconciled. Then, rather than expressly engaging in an exploration of the relative strengths and weaknesses of the two methodologies, the court simply concluded that the school districts, taking all findings into account, had met their burden of proof. Perhaps most troubling, the court decision was based on a substantial misinterpretation of the testimony of the defendant's expert. The unintentional reinterpretations of this testimony by the trial court and then the appellate court further amplified the difficulties inherent in the entire exercise.

Using these failures as context, we now explore some options for increasing the courts' likelihood of competently evaluating expert testimony—all based on tools already (but infrequently) in use. Each option looks to social scientists to play key roles in helping courts understand the testimony.

These options recognize that there exist no silver-bullet solutions; each approach carries weaknesses and risks. Yet each alternative is grounded in our belief that the key to courts' beneficial and reliable use of expert testimony is ensuring the capacity and likelihood for the testimony to be met with a competent evaluation. As Black et al. (1994, p. 743) asserted, "the real difference in scientific evidence cases is not general acceptance versus relevance/reliability, but whether or not the court is willing to undertake a thorough and active review."

Accordingly, no change in the standards of admissibility will really address the problem. The rules derived from *Daubert*, *Joiner*, *Kuhmo*, and Rules 702 and 403 are not at fault. They set forth reasonable standards and procedures for protecting juries from misleading or confusing expert testimony. But, in cases presenting complex analyses, these rules depend upon judges who are extraordinarily skilled and informed. In bench trials, too, constructive use of expert testimony depends on such judges. Unfortunately, when faced with testimony such as described in this article, few judges can or will successfully sift through the testimony to understand, for example, the paradox embedded in Approach A. We therefore see a need for additional judicial supports, focused on developing a deeper comprehension of each expert's analysis and testimony. Further, as suggested at the outset of this article, we contend that non-judicial policy makers would benefit from similar supports.

In this section, we offer seven options that may assist in addressing the problem of making complex expert witness testimony understandable to judges: (a) relaxing witness recall rules, (b) expert post-trial summaries on critical issues, (c) formal admission of social science texts, (d) pre-trial expert colloquia, (e) court-appointed scientific or technical advisors, (f) court-appointed special masters or magistrates, and (g) court-appointed supplementary experts. Each of these options has strengths and limitations.

Judicial supports should be designed to place judges in a position whereby they or their representatives can meaningfully consider the reliability of expert testimony. For instance, Rebell and Block (1982, p. 209) suggest that judicial fact-finding could be improved by relaxing witness recall rules to allow experts to rehabilitate their own prior testimony after listening to testi-

mony from opposing experts. They also suggest that experts be invited to submit post-trial summaries on critical issues. And, finally, they propose encouraging formal admission of “basic social science texts,” subject to explanatory comments from attorneys and expert witnesses. The first two options, in particular, embrace the adversarial process, offering possibilities for enhancing its effectiveness.

However, in situations where there exists a substantial disparity among the parties’ relative abilities to present a strong adversarial case (i.e., where one party’s attorneys and experts are substantially more skilled than the other’s), didactic approaches may present more effective ways to ensure judicial understanding. That is, if the goal is for the judge in such cases to successfully evaluate expert testimony, the procedures need to shift from adversarial to instructive. A court can, for instance, order pre-trial colloquia involving the parties’ experts, the court, and even neutral experts serving in an advisory capacity (Rule 16, Federal Rules of Civil Procedure). This sort of colloquium can help the court isolate and understand differences in experts’ opinions (see *In re Richardson-Merrell, Inc. “Bendectin” Prods. Liab. Litig.*, 1985, p. 1216).

In bench trials, where the judge sits as the trier of fact, this colloquium process will assist the judge in critically analyzing the testimony. A similar critical analysis by the judge is needed in jury trials, where the judge is called upon to screen out testimony that is too misleading or confusing. Another didactic option allows the court to appoint scientific or technical advisors. These advisors do not testify and cannot be deposed by counsel for the parties, and the approach therefore puts many traditionalists on edge. “I conclude that the judicial tutor or informal advisor represents an extreme departure from adversary procedures and that this form of expert appointment should be eliminated” (Deason, 1998, p. 64). Courts can, however, minimize the danger of outside-the-official-record judicial decisions by having a court reporter prepare official transcripts whenever judges converse with advisers (Black et al., 1994, p. 795; see also *Reilly v. United States*, 1988).

Alternatively, the court may appoint expert witnesses, who can testify at trial on the merits and must be available to the parties for pre-trial depositions [see Rule 706(a), Federal Rules of

Evidence]. In some cases, courts have chosen to appoint an entire panel of experts, allowing the court to gain the impressions and knowledge of a larger group (see, e.g., *Gates v. United States*, 1983, p. 1144, upholding the trial court’s dismissal of a case based on the recommendation of a panel of medical experts).

Finally, the court may appoint a special master or a magistrate. Rule 53(b), Federal Rules of Civil Procedure, authorizes a special master only to assist with complicated issues, such as is the case with difficult expert testimony. The master hears the expert testimony then submits a report to the court that becomes evidence at trial, subject to the parties’ right to object. The master may be a lawyer, a magistrate, or a scientist (see Farrell, 1994; see also Black et al., 1994). A magistrate, perhaps also serving as a special master, can make factual determinations based on a presentation of evidence. However, the magistrate’s report would not normally come into evidence; the proposed findings are merely given to the judge to use as she wishes.

The scope of the proposed testimony, as well as that testimony’s complexity, will determine its manageability. For cases presenting brief statistical analyses like those in the desegregation cases discussed earlier, the scope is relatively narrow (just two competing, concise analyses addressing a single issue)—but the content of the testimony is too complex for most judges to comprehend without assistance. Moreover, the complexity of statistical analyses can increase far beyond that presented in this desegregation litigation. Ideally, the adversarial process will shed enough light on such complexities for a judge to successfully evaluate the testimony. If not, a technical advisor or a court-appointed expert witness (discussed below) could suffice, or perhaps a pre-trial colloquium. Cases presenting a more unwieldy set of complex issues may call for a special master or magistrate. In considering which, if any, of these supports to use, a judge will hopefully be aware of her own limitations—in terms both of time available to learn about and scrutinize the testimony and of underlying knowledge and ability to understand. Unfortunately, in situations where a judge believes confusing testimony to be straightforward and understandable, she is not likely to turn to such supports.

Perhaps most controversially, we urge courts to preemptively institute procedures designed to

yield less partisan expert testimony. We favor an expansion of the judges' role in the actual appointment of supplementary experts (in addition to those chosen by the parties) in cases where the expert testimony is likely to be so complicated as to be confusing. Court-appointed experts can serve as statistical watchdogs, helping the judge to understand and interpret evidence. It is true that such a procedure may neutralize some of the adversarial 'usefulness' of experts. However, it would also, we believe, induce testimony designed to explain and educate in a similar way as would the testimony of a special master. Of course, introducing such experts into a process that is otherwise governed by adversarial rules can lead to an unhealthy reliance by the judge or jury on those labeled 'neutral'—based on a false sense that the court-appointed expert presents the only reliable testimony. (In jury trials, this concern can be partially addressed with jury instructions.) One commentator also offered the following recommendations, with which we agree:

I suggest that the informal searches for experts that judges typically conduct through their personal networks be replaced with a process that involves the parties and applies a more nuanced understanding of "neutrality." I also propose limitations on *ex parte* communications between the expert and both the judge and the parties. (Deason, 1998, p. 64.)

Of course, neither instructions, nor appointment procedures, nor communications limitations will fully address the dangers associated with judicial appointment of an expert. As with all options, there remain clear trade-offs. But the imperfections seem preferable to a lack of judicial comprehension.

Conclusion

The new No Child Left Behind Act calls for federal research dollars to be targeted toward "scientifically based research." This shift is driving a greater reliance on statistical analyses in the development of educational policy. As this article points out, this path is not without its dangers.

The courts discussed above offer an important case study. Courts considering education rights claims are increasingly presented with arguments that hinge on statistical proof of discrimination or nondiscrimination. Such testimony

often serves the beneficial role of introducing and explaining crucial and meaningful evidence. But it can also serve to obscure and mislead. Although we do not know how often these courts' efforts to understand expert testimony fall short of the mark, we do know that the fairness and outcome of such trials and hearings can turn on these efforts.

Notes

¹ The court testimony was presented in cases concerning the school districts in Rockford, Illinois and Wilmington, Delaware. Similar testimony was presented in reports and depositions in a case concerning the San Jose Unified School District, in California, where the particular dispute was settled following the expert depositions. For an expanded discussion of the San Jose case, as well as of the Rockford and Wilmington cases, see Welner (2001).

² Note that Simpson's Paradox is now explained in a publication on statistics for judges. See Kaye and Freedman, 2000, pp. 109–110 and footnote 92 (using college admissions data comparing men and women as the example).

³ See the unpublished opinion in *People Who Care v. Rockford School District*, 2000. The judge concluded that the within-school segregation evidenced in high school classes was unacceptable and must be reformed. But this decision was reversed by an appellate decision that, pointing to open enrollment, essentially absolved ability grouping as a wrongful and harmful policy, instead attributing unequal outcomes to such factors as poverty, family size, "parental attitudes and behavior . . . and ethnic culture" (*People Who Care v. Rockford*, 2001, p. 1076). The appellate decision did not address the expert witness testimony. See Note 5 for a more detailed description of the first judge's decision.

⁴ In fact, pursuant to Rule 703 of the Federal Rules of Evidence, experts can base their opinions on third party hearsay. A hearsay statement is one made by a person other than the witness on the stand, yet offered to prove the truth of the matter stated. For example, a troubling hearsay statement might be David testifying, "John told me that Mary stole the television." This testimony is troublesome because the only witness available for Mary's attorney to cross-examine is David, not John.

⁵ The school district had earlier agreed to a $+/- 12\%$ minority enrollment standard for course racial balance, and the court had so ordered. The court therefore criticized Approach A because it was based on the presumption that placements could fairly be made pursuant to test scores rather than in accord with a good

faith effort to comply with the agreed-upon standard. The court also stated that it was “swayed by” Approach B, based in part upon the relative qualifications of the experts. Perhaps most importantly, the court correctly recognized that Approach A was inconsistent with Approach B. (See the unpublished opinion in *People Who Care v. Rockford School District*, 2000, p. 41.)

References

- Appleton, D. R., French, J. M., & Vanderpump, M. P. J. (1996). Ignoring a covariate: An example of Simpson’s Paradox. *American Statistician*, 50, 340–341.
- Bickel, P. J., Hammel, E. A., & O’Connell, J. W. (1975). Sex bias in graduate admissions: Data from Berkeley. *Science*, 187, 398–404.
- Black, B., Ayala, F. J., & Saffran-Brinks, C. (1994). Science and the law in the wake of Daubert: A new search for scientific knowledge. *Texas Law Review*, 72, 715.
- Brown, H. (1999). Eight gates for expert witnesses. *Houston Law Review*, 36, 743.
- Brown v. Board of Education*, 347 U.S. 473 (1954).
- Camilli, G. & Bulkeley, K. (2001). Critique of “An evaluation of the Florida A-Plus Accountability and School Choice Program.” *Education Policy Analysis Archives*, 9(7). Retrieved June 25, 2004 from <http://epaa.asu.edu/epaa/v9n7/>.
- Coalition to Save our Children v. State Board of Education*, 901 F. Supp. 784 (D. Del., 1995), *aff’d* 90 F.3d 752 (3rd Cir. 1996).
- Coalition to Save our Children v. State Board of Education*, 90 F.3d 752 (3rd Cir. 1996).
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).
- Daubert v. Merrill Dow Pharmaceuticals, Inc.*, 43 F.3d 1311 (9th Cir. 1995).
- Deason, E. E. (1998). Court-appointed expert witnesses: Scientific positivism meets bias and deference. *Oregon Law Review*, 77, 59.
- Farrell, M. G. (1994). Scientific and technological evidence: Coping with scientific evidence: the use of special masters. *Emory Law Journal*, 43, 927.
- Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).
- Gates v. United States*, 707 F.2d 1141 (10th Cir. 1983).
- General Electric Co. v. Joiner*, 522 U.S. 136 (1997).
- Greene, J. P. (2001). *An Evaluation of the Florida A-Plus Accountability and School Choice Program*. New York: The Manhattan Institute.
- Grutter v. Bollinger*, 539 U.S. 306 (2003).
- Guillory v. Domtar Indus., Inc.*, 95 F.3d 1320 (5th Cir. 1996).
- Hand, L. (1901). Historical and practical considerations regarding expert testimony. *Harvard Law Review*, 15, 40.
- Henry, T. (2001, February 16). Florida schools snare up amid voucher threat; Findings could boost Bush’s national plan. *USA Today*, p. 1A.
- Imwinkelried, E. J. (2000). Evaluating the reliability of nonscientific expert testimony: A partial answer to the questions left unresolved by *Kumho Tire Co. v. Carmichael*. *Maine Law Review*, 52, 19.
- In re Paoli R. R. Yard PCB Litig.*, 35 F.3d 717 (3d Cir. 1994).
- In re Richardson-Merrell, Inc. “Bendectin” Prods. Liab. Litig.*, 624 F. Supp. 1212 (S.D. Ohio 1985), *aff’d*, 857 F.2d 290 (6th Cir. 1988).
- International Brotherhood of Teamsters v. United States*, 431 U.S. 324 (1977).
- Kaye, D. H. & Freedman, D. A. (2000). Reference guide on statistics, in *Reference manual on scientific evidence*, 2d ed. Federal Judicial Center.
- Kuhmo Tire Co. v. Carmichael*, 526 U.S. 137 (1999).
- Kupermintz, H. (2001). The effects of vouchers on school improvement: Another look at the Florida data. *Education Policy Analysis Archives*, 9(8). Retrieved June 26, 2004 from <http://epaa.asu.edu/epaa/v9n8/>.
- Levin, B. & Hawley, W. (1975). Forward. *Law & Contemporary Problems*, 39, 1.
- Mantel, N. (1982). Letter. *The American statistician*, 36, 46–48.
- Mickelson, R. (2001). Subverting Swann: First- and second-generation segregation in Charlotte-Mecklenburg schools. *American Educational Research Journal*, 38, 215–252.
- Oakes, J. (1995). *Ability grouping, Tracking and within-school segregation in New Castle County Schools*. Report prepared for the court in *Coalition to Save our Children v. State Board of Education, et al.* (D. Del.).
- Oakes, J. (2000). Report prepared for the court in *People Who Care v. Rockford Board of Education School District No. 205* (N.D. Ill.).
- People Who Care v. Rockford Board of Education School District No. 205*, 851 F.Supp. 905 (N.D. Ill. 1994).
- People Who Care v. Rockford Board of Education School District No. 205*, No. 89 C 20168 (N.D.Ill., August 11, 2000).
- People Who Care v. Rockford Board of Education School District No. 205*, 246 F.3d 1073 (7th Cir. 2001).
- Rebell, M. A., & Block, A. R. (1982). *Educational policy making and the courts*. University of Chicago Press.
- Reilly v. United States*, 863 F.2d 149 (1st Cir. 1988).
- Rossell, C. (1994). Report of defendants’ expert in *Coalition to Save our Children v. State Board of Education*, dated November 30, 1994.
- Rossell, C. (1999). Report of defendant’s expert in *People Who Care v. Rockford Board of Education*, dated December 10, 1999.

- Schemo, D. J. (2001, February 16). Threat of vouchers motivates schools to improve, study says. *New York Times*.
- Simpson, E. H. (1951). The interpretation of interaction contingency tables. *Journal of the Royal Statistical Society, Series B, 13*, 238–241.
- United States v. Amaral*, 488 F.2d 1148 (9th Cir. 1973).
- United States v. Baller*, 519 F.2d 463 (4th Cir. 1975).
- United States v. Hicks*, 103 F.3d 837 (9th Cir. 1996).
- Wainer, H. (1986). Minority contributions to the SAT score turnaround: An example of Simpson's Paradox. *Journal of Educational Statistics, 11*(4), 239–244.
- Welner, K. G. (2001). *Legal rights, local wrongs: When community control collides with educational equity*. Albany, New York: SUNY Press.
- Welner, K. G., & Oakes, J. (1996). (Li)Ability grouping: The new susceptibility of school tracking systems

to legal challenges. *Harvard Educational Review, 66*(3), 451–470.

Authors

KEVIN WELNER is Assistant Professor of Education, University of Colorado at Boulder, Campus Box 249, Boulder, CO 80309-0249; kevin.welner@colorado.edu. His areas of specialization are education policy, program evaluation, and education law.

HAGGAI KUPERMINTZ is Assistant Professor, Faculty of Education, University of Haifa, Mount Carmel, Haifa 31905, Israel; kuperh@construct.haifa.ac.il. His areas of specialization are assessment, methodology and cognitive science.

Manuscript Received August 27, 2002

Revision Received March 22, 2004

Accepted May 5, 2004