Testing and the Law*

Roger T. Lennon

If I had been invited to deliver a keynote address for this or any similar organization twenty-five years ago, it would never have occurred to anyone to consider as a topic the one we are addressing today. No textbook in measurement written twenty-five years ago included a chapter on legal issues in testing, and, in the unlikely event that it occurred to the author to include such a chapter, he would have been hard put to know what to cover in it. One can review the proceedings of, let us say, the ETS Invitational Measurement Conferences of the 1950's without finding any illusion to test-related legal concerns; and the same may be said of the professional literature of that era. Yet, here we are, featuring legal issues in testing as a major focus of this meeting; measurement textbooks have in fact begun to incorporate sections on testing and the law; the 1977 ETS Invitational Conference was devoted in its entirety to "Educational Measurement and the Law"; during the past decade, I have been asked to speak more often on this matter than on any other subject. Such is the change in the climate in which measurement now takes place and is perceived by society. It is now, if not commonplace, at least not at all unusual for measurement experts, whether in the academic community, in the school setting, or on the staffs of test-making organizations, to find themselves involved, willingly or otherwise, as expert witnesses in litigation having to do with testing matters.

Sources of Legal Concern

It is easy enough to discern some of the influences that have brought us to this condition. The Civil Rights Act of 1964, with its provisions seeking to eliminate discriminatory treatment of minority groups, provided the basis for statutory attacks on tests and testing practices which were shown to have a disparate impact on minority subjects. P. L. 94-142, the Magna Carta for education of the handicapped, incorporated stipulations with respect to modes of assessment and diagnosis of handicapped children, development of individualized educational programs for them, and evaluation of their progress, and outlined due-process safeguards for the observance of the conditions specified, that have together spawned legal challenges to assessment, placement, and evaluation activities relating to such children. More recently minimum-competency testing programs, to which many states have looked to insure the attainment of reasonable standards of achievement as a condition for receipt of a high-school diploma, have faced legal challenges. Increased sensitivity to concerns about invasion of privacy have been reflected in some litigation objecting to certain uses of tests. The Brown decision in 1954, calling for an end to school segration, has given rise to a series of cases in which test data and their interpretation have been central to the arguments of both plaintiffs and defendant school districts. It may also be true that we are simply becoming a more litigious society, and testing is no more immune than any other activity. And it may be the case that the widening legal involvement of testing is simply another witness to the importance that society is attaching to testing and the consequences that flow from it.

---

*Keynote address, North Carolina Association for Research in Education Annual Meeting, University of North Carolina, Greensboro, N. C., November 5, 1981 and Florida Educational Research Association Annual Meeting, Jacksonville, Florida, November 14, 1981.

Whatever the potency of these varied sources, and their relative importance in the generation of legal attention, it is clear that the measurement person, whether theoretician or practitioner, must be aware of an added dimension in the field and equip himself or herself to deal with it. The relevant professional associations have been taking steps to meet the needs of their members for this additional type of understanding, through typical activities such as establishment of committees, incorporation of appropriate material in their meetings and publications, etc.–just as we are doing here today.

The ways in which testing has intersected the law, whether in the form of litigation or legislation, have become so numerous and diverse that it is hard to know which to choose to dwell on or how to organize the domain. Within the time at our disposal today, I have elected to address a small number of areas within this field, basing my choices on what seem to me the currently most agitated issues, those with unique measurement dimensions, and those which I believe persons such as yourselves will find most relevant. I propose to address the types of problems that have arisen in connection with minimum-competency test programs; questions of test fairness, as they relate to classification of pupils as mentally retarded; due-process considerations that arise in connection with pupil evaluation, especially under P. L. 94-142; and questions of interpretation of test data submitted as evidence. Finally, I will address the types of questions posed by the so-called truth-in-testing legislation of which we have heard so much in the past two years. I will forego discussion of employment testing litigation, such as has proliferated under E. E. O. C. initiatives; issues in use of tests for selection of teachers and school administrators, though I know many of you are interested in that area; and issues involving confidentiality of test data.

I am sure you need no disclaimer from me as to any comprehensive, much less expert in the legal sense, knowledge of the total field, though I have been, for the past twenty years, something more than an amateur observer of the scene.

Minimum-Competency Testing Program Legal Issues

Public dissatisfaction with the quality of education, and particularly a sense that high-school diplomas were being awarded to students deplorably weak in the basic skills, crested about five years ago. In about two-thirds of the states, decisions were taken either by the state legislature or by the state educational agency, to mount minimum-competency testing programs, and to require as a condition for the award of a high-school diploma a stipulated level of achievement on the competency tests. Definitions of competency varied considerably from state to state, as did, correspondingly, the nature of the competency tests proposed, and as did, finally, the proficiency standards established, as reflected in the cut-off or minimum passing scores on the tests. While recognition of the questions of educational and social policy, of technical measurement issues, of administrative difficulties, and of political complications, was by no means lacking, the pressure to "do something" about declining educational standards was irresistible, and many states moved briskly ahead with the program.

Among the briskest was Florida, where a statute was enacted in 1976 which required that a student demonstrate proficiency in what was termed (infelicitously, as soon became clear) "functional literacy" in order to receive a standard high-school diploma; those not manifesting this proficiency, but meeting all other graduation requirements, were to receive a "certification of completion." A competency examination was quickly developed and had its first administration in October of 1977, though it was understood that results on the first

administration would not be used for diploma-granting purposes. In this first administration, seventy-seven per cent of black students failed the test compared to twenty-four per cent of white students. By the time the diploma requirement was to take effect, which was with respect to the graduating class of 1979, as a result of increased familiarity with the nature of the tests and opportunities for repeated taking of the tests, it developed that 1.9% of the white seniors would not receive diplomas, and 20% of the black seniors would not.

Given this dramatic disparate impact, it was not surprising that a legal challenge was mounted against the Florida minimum-competency program. This action was brought initially on behalf of certain students who had failed the test, and was soon broadened to become a class action on behalf of all who failed, regardless of race, or who might fail. The action basically challenged the constitutionality of the program on due-process grounds, asserting that the program was fatally lacking in fundamental fairness. Plaintiffs argued that students have a property right in graduation from high school with a standard diploma *if* they have fulfilled the requirements for graduation exclusive of the passage of the minimum competency or functional literacy examination; and that students have a liberty interest in being free of the adverse stigma associated with receiving a certificate of completion rather than a regular high-school diploma. According to the plaintiffs, the program did not provide sufficient prior notice to examinees, who ran the risk of suffering serious harm if they failed the test; secondly, the test measured things which the students had not had an opportunity to learn while in school; and, finally, the test was lacking in validity and reliability. Additionally, and understandably, plaintiffs alleged that the test was racially discriminatory and sought relief under Title I of the Civil Rights Act.

After a lengthy trial in 1979, the trial court handed down a decision which pleased and displeased both parties. The court agreed that there had been insufficient advance notice; it agreed that the program was racially discriminatory, in the sense that some affected graduating class members had, through some of the early years of their schooling, attended segregated and presumably inferior schools, declaring that it was unfair that such students be penalized for supposed deficiencies in their early schooling. The court therefore ruled that the test not be used as a diploma-witholding condition until all Florida pupils who had had any of their education in segregated schools should have graduated. The court upheld the right of the State Education Department to conduct a minimum-competency program and to continue the use of the test for such other non-diploma granting purposes as it might see fit; it substantially upheld the content and construct validity and reliability of the test and its freedom from racial bias.

The case was appealed to the Circuit Court, which agreed with the finding concerning advance notice. However, and here we come to the aspects of special measurement interest, the Circuit Court rejected the trial court's finding that the test was valid as "clearly erroneous." The Court declared that fairness required a showing not just of "content validity" but of "curricular validity," by which it meant a showing of a close match between test items and the curriculum to which the students had been exposed. The Court remanded the case to the trial court, asking it to have the State Education Department provide evidence that the knowledges and skills required to deal successfully with the test were ones on which the examinees had actually received instruction in the course of schooling. The appellate court declared that a state may not deprive its high school seniors of the economic and educational benefits of a high school diploma until it has demonstrated that the (minimum-competency) test is a fair test of that which is taught in its classrooms and that the racially discriminatory impact (of the test) is not due to educational deprivation.

The difficulties of accomplishing what the courts called for are easily imagined. It is not enough, for example, to show that the knowledges and skills are called for in a state course of study or even that they are reflected in state-adopted textbooks. The court is saying that there must be more persuasive showing that in the schools attended by the examinee, there was specific and appropriate instruction with respect to the tested abilities.

From a measurement standpoint, we are faced with questions of varying concepts of test validity in relation to various uses of tests. To illustrate the complexity of the technical issues, let me mention only that two weeks ago, I attended a two-day conference called precisely to consider the issues presented by this case, and to develop guidelines that would help an education agency demonstrate to a court's satisfaction the match between test and instruction, and that the panel of twenty-five or more presumed experts finished the two days' deliberation in a state that I considered one of frustration and confusion.

While the Florida case, Debra P. *vs.* Turlington, is the first and most celebrated action in this domain, it has already been joined by a similar action in Georgia. One of the county systems in Georgia, also in 1976, decreed that students had to demonstrate a stipulated level of performance on a competency test as a condition for graduation. Here, however, the board chose to use a nationally standardized norm-referenced test, the *California Achievement Test*, with the board decreeing that for diploma purposes examinees had to achieve a score of 9.0 in mathematics and reading. As in Florida, a disproportionate number of those who failed to meet the standard were black and, also as in Florida, several legal challenges were mounted against the test diploma requirement. The heart of the challenge again was violation of equal protection, since blacks had had inferior educational opportunities, and violation of due process, both because of inadequate advance notice, and because of poor fit between the *California Achievement Tests* and the actual instructional program in the schools.

The court rejected the plaintiffs' inadequate-notice challenge and endorsed the decision of the school board to use a nationally-normed, well-established test rather than what it called–harshly, I thought–Florida's "hastily concocted examination." The court held, however, that the diploma withholding on the basis of inadequate test performance was unconstitutional because the school system had not shown that the test material was actually taught in its schools. The school system presented, as an expert witness, a California Test Bureau staff member who demonstrated the match between the test content and the content of the Schoot-Foresman arithmetics used throughout the county schools, but the court found this expert testimony unpersuasive.

I think it is fair to say that the legal situation with respect to the ability of a test to withstand judicial scrutiny as a basis for diploma granting, is unclear. Efforts are being made, I understand, to arrive at some non-judicial ways of resolution in Florida. In the meantime, as measurement people, we are being forced to review traditional concepts of content validity, to attend to something called "instructional validity"–a much less familiar term in our professional literature–and, in the process, finding ourselves, at least by my analysis, back at construct validity of achievement tests. It is interesting how philosophical differences as to whether minimum-competency tests for graduation should focus on the so-called life skills, or survival skills, *versus* concentrating on mastery of what has actually been taught, rather strictly construed, surfaces in the seemingly technical validity debate.

Culture Bias in Intelligence Tests

Let me turn now to problems of test fairness for classification purposes as they have become legal matters. Most famous of the cases, I believe, is Larry P. *vs.* Riles. Larry P. *vs.* Riles is an action first brought in November 1971 on behalf of a black child (pseudonym "Larry P.") against the San Francisco Unified School District and the California State Education Department, challenging as unconstitutional the use of standardized intelligence tests for placement of black children in classes for educable mentally retarded in the San Francisco schools, citing disproportionate representation of blacks in such classes as *prima facie* evidence of discrimination. Judge Robert Peckham granted a preliminary injunction in favor of the plaintiffs which restrained the San Francisco School District and the California State Education Department "from placing black students in classes for the educable mentally retarded on the basis of criteria which place primary emphasis on the results of I. Q. tests...if the consequences of use of such criteria is racial imbalance in the composition of such classes." Defendants appealed this preliminary injunction, but it was affirmed by the Court of Appeals in 1974.

Plaintiffs then moved to expand the class to include "all black California school children who have been or may in the future be classified as mentally retarded on the basis of I. Q tests," and likewise to extend the terms of the preliminary injunction to restrain defendants from "performing psychological evaluation of...black California school children by the use of standardized individual ability or intelligence tests which do not properly account for the cultural background and experiences of these children." In January 1975, the State Education Department voluntarily extended the moratorium on I. Q. testing for E. M. R. placement to all California children regardless of race.

The plaintiffs further amended their complaint to allege several additional statutory bases. These additional elements included violation of Title VI of the Civil Rights Act of 1964 and of the Emergency School Aid Act of 1972 and 1974. The United States Department of Justice entered the case as amicus curiae in August 1977, adding to the complaint alleged violations of the Education of All Handicapped Children Act and the Rehabilitation Act of 1973.

The case went to trial in October 1977. Thetrial was a lengthy one, generating a transcript of more than 10,000 pages. Much of the testimony was that of expert witnesses called upon by both plaintiffs and defendants to testify concerning the nature of intelligence tests, their uses and limitations, and their bias or lack of it.

Following are excerpts from the Peckham decision that represent its essential findings:

> "This court finds in favor of plaintiffs...defendants have utilized standardized intelligence tests that are racially and culturally biased, have a discriminatory impact against black children, and have not been validated for the purpose of essentially permanent placements of black children into educationally dead-end, isolated, and stigmatizing classes for the so-called educable mentally retarded....

"The unjustified toleration of disproportionate enrollments of black children in E. M. R. classes, and the use of placement mechanisms, particularly the I. Q. tests, that perpetuate those disproportions, provide a sufficient basis for relief under the California Constitution... and under the federal Constitution....

"Defendants' conduct, in connection with the history of I. Q. testing and special education in California, reveals an unlawful segregative intent.... This intent, consistent only with an impermissible and unsupportable assumption of a higher incidence of mental retardation among blacks, cannot be allowed in the face of the constitutional prohibition of racial discrimination...."

The Peckham decision prescribes certain remedies, which include (1) making permanent the original temporary injunction prohibiting the use of any standardized intelligence test for the identification of black educable mentally retarded children, or their placement into educable mentall retarded classes, without prior approval of the court; (2) requiring defendants to monitor and eliminate disproportionate placement of black children in California's educable mentally retarded classes through procedures prescribed in the decision; (3) requiring the defendants to re-evaluate every black child currently identified as an educable mentally retarded pupil without reference to or dependence on results of any standardized intelligence test, and to prepare an individual educational plan for each such child following the re-evaluation, which plan will specify the type of supplemental assistance needed to allow the child to return to a regular classroom when it is judged that he has been mis-classified in the first instance.

Assignment to E. M. R. classes, far from being seen as an attempt to provide a suitable, more beneficial type of educational program, is throughout the judge's decision seen as virtually a prison sentence. Classes for the mentally retarded are repeatedly characterized as "dead-end classes," with instruction that "deemphasizes academic skills," in which pupils "inevitably lag farther and farther behind the children in regular classes" and as classes that doom the pupils to stigma, inadequate education and failure to develop the skills necessary to productive success. Peckham endorses completely the plaintiff's allegations concerning the hurtful consequences of assignment to classes for the mentall retarded.

The concept that the educational treatment proffered in classes for the mentally retarded was as poor as suggested above is at the heart of the plaintiff's case. Had E. M. R. classes been seen as providing beneficial and appropriate treatment, it is possible that the test issue would not have arisen, though the plaintiff's strategy clearly was to make the tests the central target.

By far the most important part of the Peckham decision from a measurement standpoint is his discussion of the nature, purposes and uses of intelligence tests and the question of their cultural bias. As noted earlier, the trial featured the testimony of many measurement experts who testified about the history of intelligence testing, the nature and purposes of such testing, the cultural bias issue, and the relevance of intelligence test scores in the classification of children as mentally retarded. Judge Peckham's consideration of the bias issue begins with the universally recognized systematic differences in the scores of black and

white examinees on almost all such tests. He then proceeds to consider numerous possible explanations for this difference, including genetic reasons, environmental reasons, and test-related reasons, and decides overwhelmingly to accept the test-bias explanation rather than any of the others.

The possibility of any genetic or constitutional differences between blacks and whites that would account for their different performance on I. Q. tests gets short shrift: the judge declares this position to be insupportable, or even worse, impermissible. Nor is he much more sympathetic to the defendants' position that the inferior black performance on I. Q. tests may be ascribed to their lower socio-economic and generally disadvantaged status; he did not find persuasive testimony to support the notion that such impoverishment leads in any way to mental retardation. He concludes, therefore, that the most sensible explanation for the observed differences is that the tests simply put children from the black culture at a disadvantage and thus give rise to spurious estimations of their lack of mental ability.

The judge declared that there is very little information concerning the validity of intelligence tests, either group or individual, for black pupils. He is insistent (in the face of a now strong consensus to the contrary) on seeking evidence of differential validity, declaring that the findings with respect to discriminatory content make it easy to believe that the tests will be less valid for blacks than for whites.

We would say that the judge's opinion failed to take account of the complexity of the bias issue; for example, there is no recognition of the multiple definitions of "bias" that pervade psychometric literature; that there was an axiomatic assumption that there are no true inter-group differences in the attributes measured by tests such as WISC; a disinclination to consider other information about subsequent careers of *Larry P.* and the other cases at issue which strongly suggests that they were, in fact, mentally retarded; and a slighting of evidence suggesting that poor WISC performance may be ascribable to environmental deprivations. The judge came down unequivocally and solidly on the side of those who find the explanation for black-white differences in the instrument and not in the groups.

It became appropriate in the trial to consider whether the assignment processes used after the initial injunction against the use of I. Q.'s were effective. Testimony was given to the effect that the non-test-based processes seemed to be satisfactory, but an interesting note (and a circumstance clearly a little embarrassing to Peckham) is that the proportion of blacks assigned to classes for the mentally retarded since the banning of I. Q.'s as a basis for classification remained at just about the same level as before the ban–and this in spite of a California legislative mandate that the representation of various groups in classes for mentallyretarded should not be disproportionate.

It would be easy–and useful–to devote several hours to a detailed consideration of the Peckham argument, and to evaluate the decision in the light of relevant research. We don't have the time for that, but fortunately, the job has been done–very well, in my judgment–by Nadine Lambert, a school psychologist who was one of the defense witnesses. Her critique appears in the September 1981 issue of *The American Psychologist*, and I commend it to you.

Soon after the Peckham decision was handed down, I wrote, in a memo to our staff,

"The decision has been hailed as a sweeping attack on I. Q. tests. It is not. In fact, Peckham declares that 'our decision in this case should not be construed as a final judgment on the scientific validity of intelligence tests.' It is important to realize that the supposed 'findings of fact' are those deriving from a single trial, in a single jurisdiction, in light of a single judge's evaluation. There is no assurance that judges in other circuits, confronted with the same issue, and the same 'facts,' would come to the same conclusion as Judge Peckham; nor is it certain that even Peckham's evaluation will be sustained on appeal."

As if to illustrate this point, within three months of Larry P. there was handed down a decision by a federal court of Illinois in a remarkably similar case, precisely contradictory of the Larry P. ruling. This Illinois case, known as PASE *vs.* Hannon (PASE meaning Parents in Action on Special Education) was brought on behalf of black students in the Chicago school system who protested their assignment to classes for mentally retared on grounds of allegedly biased testing–chiefly WISC and Stanford-Binet. Trial covered much the same ground as Larry P., and heard from many of the same expert witnesses. Judge Grady's ruling was to the point:

"I believe, and today hold, that the WISC, WISC-R and Stanford-Binet, when used in conjunction with the statutorily mandated other criteria for determining an educational program for a child, do not discriminate against black children in the Chicago schools."

Judge Grady's approach was singular, to say the least. He took a dim view of the testimony of experts for both sides, commenting that most of them seemed more intent on promoting their doctrinaire views on the issues than on enlightening the court on their scientific merit; he even expressed doubt that many of the experts had ever seen the tests in question. This being the case, said the judge, he knew no better way to proceed than to examine the tests himself, item-by-item, and form his own opinion as to their fairness, or lack of it. This he proceeded to do, reading into the record (to the horror of the copyright owners) every item in WISC-R and Stanford-Binet, with his own evaluation of their freedom from bias. He concluded that only a handful of the more than 300 items might be biased, and therefore ruled as he did.

Few measurement people will applaud Judge Grady's approach to test validation or bias; but until the professionals have their act together, we should not be surprised to see other judges behaving in similar fashion.

And, to come even closer to home, I refer to an action brought in Hendry County in Florida last year on behalf of several black plaintiffs, designated S-1, S-2, S-3, etc., alleging, among other things, unfairly discriminatory evaluations for assignment to classes for mentally retarded. The complaint declares, "Plaintiffs...challenge defendants' use of racially and culturally discriminatory intelligence tests and evaluation procedures to identify and place children in special education classes for educable mentally handicapped students." I have no word as to movement of this action through the courts.

Thus, we find ourselves in a confused legal state on this matter–contradictory decisions from courts of equal status, with the California case on appeal. (The Chicago schools, ironically, have announced that they will discontinue the use of intelligence tests for E. M. R. placement, despite the Grady decision.) Our usual judicial-adversarial procedures seem ill-suited to handle technical issues such as those involved in test bias, but how else are all interests to be protected?

P. L. 94-142 and Testing

Let me now invite your attention to another area in which there are already signs of legal difficulties for testing. This is in the activities being carried out in implementation of Public Law 94-142, the Education for All Handicapped Children Act. As most of you are aware, rules and regulations have been promulgated prescribing how state and local educational agencies are to carry out the provisions of the act.

Let me quote for you from the rules and regulations having to do with evaluation procedures:

> "State and local educational agencies shall insure, at a minimum, that:
>
> (a) Tests and other evaluation materials:
>
> (1) Are provided and administered in the child's native language or other mode of communication, unless it is clearly not feasible to do so;
>
> (2) Have been validated for the specific purpose for which they are used; and
>
> (3) Are administered by trained personnel in conformance with the instructions provided by their producer;
>
> (b) Tests and other evaluation materials include those tailored to assess specific areas of educational need and not merely those which are designed to provide a single general intelligence quotient;
>
> (c) Tests are selected and administered so as best to ensure that when a test is administered to a child with impaired sensory, manual, or speaking skills, the test results accurately reflect the child's aptitude or achievement level or whatever other factors the test purports to measure, rather than reflecting the child's impaired sensory, manual, or speaking skills (except where those skills are the factors which the test purports to measure);
>
> .... (f) The child is assessed in all areas related to the suspected disability, including, where appropriate, health, vision, hearing, social and emotional status, general intelligence, academic performance, communicative status, and motor abilities. ....
>
> " (a) In interpreting evaluation data and in making placement decisions, each public agency shall:
>
> (1) Draw upon information from a variety of sources, including aptitude and achievement tests, teacher recommendations, physical condition, social or cultural background, and adaptive behavior;

(2) Insure that information obtained from all of these sources is documented and carefully considered;

(3) Insure that the placement decision is made by a group of persons, including persons knowledgeable about the child, the meaning of the evaluation data, and the placement options; and

(4) Insure that the placement decision is made in conformity with the least restrictive environment rules in §§ 121a.-550-121a.554.

(b) If a determination is made that a child is handicapped and needs special education and related services, an individualized education program must be developed for the child in accordance with §§121a.340-121a.349 of Subpart C."

These are well-intentioned regulations seeking to provide optimum assessment of handicapped children. The trouble is that if interpreted literally, they simply cannot be compiled with because tests of the kinds contemplated are simply not available. If you consider the varieties of handicapped children who are the subjects of P. L. 94-142 and consider the requirement that only tests be used with these children whose validity for them has been established, it quickly becomes clear what a gap exists between the tests needed and the tests available. As long as the regulations are viewed as declarations of an ideal toward which practitioners should strive as far as the present state of the art and available instruments permit, no harm will be done. But the tone of P. L. 94-142 is one of vigorous protection of the rights of the handicapped, a tone which it seems to me has encouraged greater litigiousness on the part of parents and other advocates of these students. The language of the regulations on evaluation procedures provides a basis for legal challenges on due process grounds to almost any testing used for diagnostic and evaluation purposes.

Interpretation of Test Data as Evidence in Litigation

In addition to cases such as those described above, in which tests are the primary target of legal challenge, there are many cases in which the interpretation of test data submitted as evidence by either plaintiffs or defendants is critical. Examples of such cases would surely include actions brought in the wake of the Brown desegregation decision in which plaintiffs have challenged the continued or contemplated conduct of racially imbalanced schools as contrary to the Brown ruling, adducing test data, both as evidence of unfairly discriminatory practices and as evidence of unequal delivery of educational services. The latter inferential chain is a common one: average differences in achievement test scores between pupils in preponderantly non-white schools and pupils in preponderantly white schools are cited by plaintiffs as documenting charges of inadequacies in the educational programs of the non-white schools. Interestingly, such reliance on test data often occurs in the same suit in which test data are attacked as giving rise to unconstitutionally discriminatory placement and classification.

In these actions, the demands on the measurement person involved as witness run to the proper interpretation of scores rather than to the nature of the tests themselves. They call for understanding of the sources of variance in achievement test scores and the extent to which they may be regarded as indices of a quality of educational effort.

Truth-in-Testing Legislation

It is not only in the courthouse that testing and the law come together. Concern with the consequences of certain types of testing, particularly those relating to admission to post-secondary educational programs, has resulted in a flurry of legislation at both the state and the federal level calling for greatly increased disclosure of the content of such examinations, of the performance of individual examinees, of scoring keys, and research data about the tests. Legislation of this kind generally proffered under the designation of truth-in-testing bills has, over the past two years, been introduced in about twenty state legislatures and in both this and a prior session of Congress; thus far only two states, New York and California, have seen fit to adopt the proposed bills. Proponents argue that when the consequences of testing are so serious for an individual, such as his being denied admission to the college of his choice, or to a law school, or a medical school, elementary justice demands that the examinee have an opportunity to review a copy of the examination and of his answers, and, more generally, that the test agency be obliged to make copies of the examinations and supporting research data available to state education agencies or the Department of Education. The test-making agencies involved in this type of test development have in general resisted this type of legislation, although one of these agencies, the College Entrance Examination Board, has on its own adopted disclosure policies that come close to meeting those that would be established legislatively. To suggest the arguments advanced by the test-making agencies, I cite from a statement which we prepared for the Committee of the Florida House of Representatives which earlier this year was considering a Florida truth-in-testing bill.

> "We accept that persons required to take examinations as a condition for entry into post-secondary institutions, or occupations, should be fully informed about the nature of the examinations and their use. We believe that the public should have access to appropriate information regarding the development and uses of such tests. We agree that examinees have bases for confidence that their tests are scored accurately, and that results are transmitted correctly to institutions. Tests ought to be used in a manner that provides fair treatment and equal rights for all examinees....
>
> "In spite of our concurrence in these avowed purposes, we oppose enactment of H. B. 0003 for the following reasons:
>
> - There has not been a substantial showing of need for legislation of this kind.
>
> - The controls and refinements of practice sought by the bill are better handled through appropriate professional organizations than through legislation. These organizations have, in fact, addressed themselves to many of the issues of test information that the proposed bill looks to.
>
> - Certain of the alleged 'abuses' in admissions testing cited in the bill's preamble result from practices of user institutions, and not of the testing agencies which are the targets of the bill.

- The proposed bill will *not* achieve a heightened level of knowledge and understanding on the part of examinees and the general public.
- The proposed bill will inevitably lead to higher costs and diminished services to examinees, without offsetting benefits."

While truth-in-testing bills have been by far the most common of the attempts to legislate testing, they have not been the only ones: bills seeking either to ban testing of various kinds, or to mandate testing of various kinds, keep appearing, and it is up to the measurement community in the jurisdictions affected to stay alert to the introduction of such bills and to make informed views known concerning them.

The Meaning for Measurement

It is impossible not to wonder about the significance for measurement of these mounting involvements with lawyers and legislators. Is it a good thing for us measurement people to have to ply our trade in the courtroom and in the legislative chamber? I have voiced some of my reservations about the appropriateness of adversarial judicial procedures for the resolution of scientific and professional problems. I have begrudged the hours, days, and weeks that I have had to deveote to proceedings in which I have been involved in one capacity or another. I have not found it comfortable to have some of my deepest professional values and practices called into legal question, and even rejected. Yet, when all is said and done, I believe that involvement in litigation and in legislation is more helpful than hurtful to measurement. It is in a sense an ultimate validation of the social utility, if not of the scientific merit of what we do. It is witness to the fact that tests and their uses are now seen by society as critical–so critical that they have properly become the object of litigation and legislation. If that is the price we have to pay for recognition of the importance of our efforts, it does not seem to me excessive.