



If It's Broke, Fix It!

Carole L. Perlman

Educational Researcher, Vol. 23, No. 3. (Apr., 1994), pp. 37-38.

Stable URL:

<http://links.jstor.org/sici?sici=0013-189X%28199404%2923%3A3%3C37%3AIIBFI%3E2.0.CO%3B2-8>

Educational Researcher is currently published by American Educational Research Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/aera.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

where the strata or classes of ability are defined in terms of generality. General intelligence is located at the top of the hierarchy, representing the finding that virtually all cognitive tests correlate positively. The middle or second-stratum level consists of broad abilities: fluid intelligence, crystallized intelligence, general memory and learning, broad visual perception, broad auditory reception, broad retrieval ability, broad cognitive speed, and processing speed. These broad abilities organize dozens of narrower, third-stratum abilities at the bottom of the hierarchy. Chapter 17 discusses Carroll's three-stratum theory of ability in relation to the literature on the

genetic basis of ability, the development and change of abilities over the life span, and the extent to which abilities can be modified. The final chapter discusses implications of his organization of the literature for the future of mental measurement.

It is obvious from this review that I think Professor Carroll has done a great service for researchers in the field of individual differences; I have no doubt his book will be read for a long time to come. Of course, he has been doing this his entire career. Perhaps more importantly, however, John Carroll has done all educational researchers a great service. The research acumen and scholar-

ly ability demonstrated in this book set a high standard for generations of future educational researchers. I would therefore encourage all educational researchers, not just those in the field of cognitive abilities, to sit down with this book for an hour or two. I think it will be time well spent.

References

- Carroll, J. (1976). Psychometric tests as cognitive tasks: A new "Structure of Intellect." In L. Resnick (Ed.), *The nature of intelligence* (pp. 27-56). Hillsdale, NJ: Erlbaum.
- Schmid, J., & Leiman, J. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53-61.

If It's Broke, Fix It!

The Fractured Marketplace for Standardized Testing. Walter M. Haney, George F. Madaus, and Robert Lyons. Boston: Kluwer, 1993, 347 pp., \$59.95. ISBN 0-7923-9338-4.

Review by CAROLE L. PERLMAN
Chicago Public Schools

Surely you know their names. You've seen the names—Educational Testing Service, The Psychological Corporation, National Computer Systems, Scantron—festooned across the exhibition booths at the AERA Annual Meeting. It is obvious from even a perfunctory visit to the exhibit area that testing is very big business indeed.

The Fractured Marketplace for Standardized Testing, which examines the testing industry and its profound impact on the U.S. educational system, evolved from papers prepared in the late 1980s for the National Commission on Testing and Public Policy. It is worth noting that one of the commission members is a former governor of Arkansas who is currently influential in shaping federal testing policy.

Despite occasional forays into personality, employment, and certification testing, the book's primary focus is on K-12 achievement tests and college admissions exams. The authors use the term "fractured marketplace" in several

different ways. The first, which is discussed only briefly, refers to their observation that different segments of the testing marketplace might be characterized as monopolies, oligopolies, and free markets. The second alludes to the large number of companies that represent at least some part of the testing market. The third usage refers to the increasing trend for tests to be sponsored, developed, administered, scored, and interpreted by different organizations. These various definitions aside, the overall theme of the book is that the marketplace is broken and seriously in need of repair.

The Nature of the Marketplace

The first part of the book gives a survey of the testing marketplace and a description of the players: a varied collection of large, medium-sized, and small firms. Some noteworthy trends in recent years include mergers and acquisitions, greater use of computer technology for test development, scoring, administration and interpretation, and the previously mentioned splintering of responsibility for a single testing program among several organizations. The authors contend that this diffusion of responsibility, along with government policies, and economic considerations have often resulted in tests being used inappropriately. One example is the use of the SAT for purposes that were never

intended, for example, to rank order states on a "wall chart" or to determine which students may participate in intercollegiate athletics.

The authors calculated estimates of how much testing takes place and how much it costs—a brave endeavor, since much sales information is either secret or impossible to tease out of the annual reports of the test publishers' parent companies. The authors maintain that besides the obvious direct costs of developing, purchasing, and scoring standardized tests, there are additional (and much greater) indirect costs, the most salient being the opportunity cost of the time that is spent on preparing for and giving the tests. "That opportunity cost is the value of the *next best alternative* to testing, which is, presumably learning, regardless of whether such learning is going on or not" (p. 103). Based on a cost-effectiveness analysis, the authors assert that "whatever the benefits of testing . . . as long as we fail to take into account the value of student and teacher time invested in [standardized] testing, we will invest far more than is optimal" (p. 124). However, one could have just as easily applied the same logic to—and probably drawn a similar conclusion about—any other form of assessment.

Readers hoping to find timely and accurate figures on the testing industry will be disappointed. One problem is

the use of old data. Despite the 1993 copyright, a great deal of the data presented are from the mid-1980s and much has changed since then. For example, the estimated number of state assessments administered in Illinois is 7,500 per year. Yet, as early as 1988, the actual number had soared to 337,920, and by 1993, the total was 815,575 (D. Goedecke, personal communication, November 5, 1993); the State Student Assessment Program Database developed by the Council of Chief State School Officers and the North Central Regional Educational Laboratory (1993) provides much more current information. A second difficulty is that some estimates are based on questionable assumptions. A case in point: The estimates of the number of district-level assessments were inexplicably derived by examining the testing done in only three districts—Topeka, Boston, and a suburban Massachusetts district—and extrapolating those findings to the rest of the country. The waters are further muddied by counting each separately scored subtest as a separate test, so that a single ACT would be counted as four tests. Despite these problems, it is hard

to argue with the authors' conclusion that much testing takes place and that the costs are considerable.

The authors devote a chapter each to an examination of the social forces that have led to increased standardized testing and an extended and illuminating exploration of several factors that serve to compromise the quality of inferences that can be drawn from test results. Among these are the Lake Wobegon phenomenon, in which an implausibly large number of school districts score above the national norm on standardized tests; the manner in which tests are advertised and marketed, with validity evidence taking a backseat to other concerns; test preparation and coaching; and the Request for Proposals process used to purchase tests and related services, which militates against conducting some necessary but expensive types of validation studies. The authors conclude that "though the problems we have identified in the testing marketplace may not represent a major blight on the entire enterprise of testing, neither do they represent merely minor mold that can be safely ignored" (p. 245).

Mending the Marketplace

What, then, to do about the fractured marketplace? The authors discard as possible remedies a greater reliance on market forces, professional standards and peer review, litigation, education of test users, government regulation, and an independent auditing agency. Will switching to performance assessments help? Not likely, say the authors, since those tests would also need to be validated for each intended use, the results are subject to corruption in a high-stakes environment, and there is evidence that they would be developed and marketed by the same companies that are major forces in the standardized testing marketplace.

The authors feel that standardized testing itself is less a problem than the inordinate weight that is placed on standardized test results:

[A] strategy useful in helping to solve the problems of the fractured marketplace . . . is to reduce the weight placed on this technology . . . and to disentangle the myriad functions (e.g., student evaluation, instructional guidance, and school and teacher evaluation) that the same standardized testing program is often expected to perform. (p. 294)

The authors recommend using a combination of different types of assessments and avoiding the tendency to try to make any single test serve too many different purposes, especially those for which validity has not been established. This, coupled with concerns regarding the validity of performance assessments, plainly bodes ill for national testing programs that seek salvation in multipurpose performance assessments.

The Fractured Marketplace for Standardized Testing has much to recommend it. The overarching emphasis on validity is both refreshing and necessary. The historical examples and discussion of legal issues provide thought-provoking reading, as well as a valuable context for interpreting current developments. This book provides a needed perspective on a topic that is too often clouded by rhetoric, politics, and expediency.

Reference

Council of Chief State School Officers & North Central Regional Educational Laboratory (1993). *State Student Assessment Program database* [Machine-readable data file]. Oak Brook, IL: North Central Regional Educational Laboratory.



University of Pittsburgh

Lilly Endowment Evaluation Fellowships

In the fall of 1990, the University of Pittsburgh and the Lilly Endowment, Inc. formed a partnership designed to enhance the Endowment's use of evaluation in its education programs. One important aspect of this partnership is the Endowment's sponsorship of Fellowships to attract and prepare new entrants to the field of evaluation, especially women and minorities. The Fellowships provide full support (tuition and stipend) for doctoral study in the Policy, Planning and Evaluation Program of the Administrative and Policy Studies Department in the School of Education at the University of Pittsburgh.

A new cohort of Lilly-supported Fellows will be selected for the fall of 1994. The program of studies emphasizes decision-oriented educational research, program documentation, evaluation and decision making, and other innovative approaches to providing information useful to decision makers at program, institution, state and national levels. Fellows will be working with senior faculty, including William Bickel, William Cooley, and Tony Eichelberger. Applicants must be interested in pursuing a career in evaluation and have a Masters Degree or approximately 30 credits of graduate study that would apply to this doctoral program.

The application should be completed as soon as possible. For a complete set of application materials, call 412-624-7090 or write to:

**Lilly Endowment Fellowships, Suite 7A, LRDC Building,
3939 O'Hara Street, Pittsburgh, PA 15260**