

## Book Reviews



---

**HANDBOOK OF TEACHER EVALUATION.** By *Jason Millman* (Ed.). Sage Publications, Beverly Hills, CA (348 pages; \$27.50), 1981.

Reviewed by: *W. J. McKeachie*  
*Professor of Psychology*  
*University of Michigan*  
*580 Union Drive*  
*Ann Arbor, MI 48109*

“In the spring of 1975, a task force of the National Council on Measurement in Education, an organization devoted to improving the practice of educational measurement, recommended that a group of writers prepare a teacher evaluation guidebook for practitioners. The proposed attributes of the work were as follows”:

The guidebook should focus on the currently available approaches for evaluating teachers, detailing the strengths and shortcomings of each. Readers should, as a consequence of completing the guidebook, be in a position to know the special advantages and distinctive liabilities of all commonly employed teacher evaluation strategies.

The guidebook should be written in practitioner’s language. The writers should be instructed to prepare a volume which should be easily comprehended by busy educators. The writers should not endeavor to impress their scholarly colleagues in the guidebook. Rather, it should serve as a guide to the practice of teacher evaluation, kindergarten through college.

Have Millman and his coauthors achieved the purposes recommended? On the whole, I judge that they have. While the chapters vary somewhat in readability and in rigor, they cover the major approaches for evaluating teachers and they do provide reasonable assessments of the strength and shortcomings of each.

It is common for reviewers of an edited book to criticize the book for lack of unity. Since it seems unlikely that many readers will wish to read this book from cover to cover, such criticism seems inappropriate in this case. Such a reader would find some duplications, some inconsisten-

cies, and some lack of continuity, but in a handbook most users will be looking up a particular topic in the hope of finding reasonable coverage of this aspect of the problem of teacher evaluation. Thus, it is not inappropriate that reliability is discussed at an introductory level in more than one place or that authors sometimes reiterate points that have been made earlier.

The *Handbook* is reasonably organized into three parts:

- I. "Orientation"
  - (1) Introduction, Jason Millman, posing the questions, "Who should evaluate?" "For what purpose?"
  - (2) Criteria of Good Teaching, Robert M.W. Travers
  - (3) Context/Environment Effects in Teacher Evaluation, Bernard H. McKenna
- II. "Sources of Evidence"
  - (4) Teacher Interviews, Donald L. Haefele
  - (5) Teacher Command of Subject Matter, William U. Harris
  - (6) Peer Review: Documentary Evidence in the Evaluation of Teaching, Grace French-Lazovik
  - (7) Classroom Observation, Carolyn M. Evertson and Freda M. Holley
  - (8) Student Ratings of Instruction, Lawrence M. Alcamoni
  - (9) Student Achievement as a Measure of Teacher Competence, Jason Millman
  - (10) Beyond Classroom Walls: Indirect Measures of Teacher Competence, Jean A. King
  - (11) Faculty Self-Evaluation, J. Gregory Carroll
- III. "Systemic Uses of Evidence"
  - (12) Contract Plans: A Professional Growth-Oriented Approach to Evaluating Teacher Performance, Edward F. Iwanicki
  - (13) Evaluation-Based Teacher Development, Stephen C. Brock
  - (14) Summative Teacher Evaluation, Michael Scriven
  - (15) Politics of Teacher Evaluation, John D. McNeil
  - (16) The Political Realities of Teacher Evaluation, Mary Louise Armiger
  - (17) *Fairness and the Legal Context of Teacher Evaluation*, Kenneth Strike and Barry Bull

While all of the authors take as a given the need for evaluation, they have a healthy skepticism about the quality of the various sources of data available. As Millman points out in Chapter 1, multiple sources of evidence are generally recommended, but more is not necessarily better. Travers in Chapter 2 poses a number of cautions for evaluators, noting that not all models of teaching would hold the teacher responsible for pupil learning and it is unlikely that any one set of criteria is appropriate for all teachers. McKenna in Chapter 3 suggests that an input-output model is useless because it fails to take adequate account of the context within which teaching and learning take place. Thus, if one reads the chapters in order, one may wonder whether it is even worthwhile to continue reading Part II, "Sources of Evidence."

Most of the chapters in Part II review the negative evidence with respect to the validity of the particular method described in that chapter and then go on to suggest ways in which the method may be appropriately used. Considering that all of the authors are active researchers in the field, I was a bit disappointed in the number of "shoulds" and "musts," usually unsupported by evidence. Nonetheless, the chapters are reasonably good representations of the wisdom the evaluators have acquired over the years with respect to each of the methods of evaluation discussed.

To those active in the field of evaluation it will come as no surprise that the most provocative chapter is Scriven's "Summative Teacher Evaluation." Scriven begins by making a distinction between "worth" and "merit." A teacher may have great merit with respect to the role that teacher performs, but that role may still not be of great value or worth to the institution. A high degree of merit thus, does not insure that one should be promoted or given tenure.

Discussing summative questions on student ratings, Scriven says that one should never use the question, "How much have you learned?" but rather "How well did the teacher teach?" To me, the ultimate test of which is the appropriate question is which provides the more valid answers. Students have stereotypes about what good teaching is and students also fall prey to the same sort of misjudgments about how teacher characteristics relate to teacher effectiveness that Scriven inveighs against with respect to faculty peers and administrators. Consequently, it seems likely that student answers to the question "How much did you learn" may be a more valid measure of teaching effectiveness than answers to the question "How well did the teacher teach?" What-

ever the question asked, one must still take into account differences in student assignments, character of courses, and so forth.

Scriven makes a good point in warning against overkill in evaluation—evaluating every course every term. The costs in terms of time of students, faculty, and evaluators are significant, and the gains are not likely to necessitate such an expenditure.

It is interesting that very few of the authors of chapters of the *Handbook* give much attention to the costs in terms of student and faculty time involved in the methods of evaluation they describe. While most of these authors have something to say about the possible strengths or weaknesses of these methods, the typical conclusion is that properly handled, some benefits may accrue from their use. Whether or not these benefits justify the time and effort is not typically addressed.

Scriven is much concerned about possible bias and comes down strongly on the necessity for an objective determination in advance of the weights of each component entering into the evaluation. He argues that the same evaluative approach should be used by all departments and schools although he admits that it may be useful politically to allow different approaches by different departments. This obviously is in direct contradiction to Travers' warning that no single set of criteria is appropriate for all teachers, let alone for all departments. (I look forward to a debate on this question the next time Travers and Scriven appear in the same symposium.)

The issue relates to the general assumption on the part of most of the authors that planning, clear specification of goals, and systematic approaches to the kinds of data and weight to be given to data are not only good, but essential, if one is to have an appropriate system of evaluation. While arguing against this point may seem analagous to arguing against motherhood, apple pie and the American way of life, I do so.

We sometimes forget that the purpose of teacher evaluation is to improve education and teaching. The empirical evidence with respect to specifying objectives, whether in business or education is not terribly encouraging. It may well be that less formal, more ambiguous procedures actually produce better results than carefully constructed systematic methods. (Unfortunately, many institutions may not have much latitude with respect to this point since, as is indicated in the chapters on political and legal aspects of teacher evaluation, we may be forced into relatively rigid systems of teacher evaluation.)

While I am probably one of the few individuals who will read the book through and would not suggest that this is the best way to approach the *Handbook*, I do believe that both practical evaluators in schools and colleges and evaluation researchers will find that dipping into the book reveals enough nuggets to justify its purchase.