Book Review


JIM HORN, PH.D.
Louisiana State University, Shreveport

Most people, including some of us educationalists, have long suspected that the organization of schools, the work of teachers, and student learning are necessarily linked. Of course, this link is readily acknowledged when high levels of student learning would seem to provide sure signs that good schooling in general and good teaching in particular are occurring. An uneasiness sometimes ensues, however, when stakeholders begin to examine how lower levels of student learning may be linked similarly to the quality of schools and the quality of teaching. Questions surely arise that challenge the commonsensical notion that a clear causal chain exists between student learning and the quality of schools and teaching. For instance, can “good teaching” in rundown inner-city schools where social and economic problems abound be as effective as “good teaching” in well-funded suburban districts where family and social structures provide nurturing environments? Or can measurement tools that meet psychometric criteria for validity and reliability be applied to gauge complex learning tasks within local contexts? Perhaps more basic still, do complex linkages exist within learning institutions that confound our efforts to establish direct cause-effect relationships among school organizations, acts of teaching, and evidence of learning?

As Millman and Schalock indicate in their introduction to Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?, these are not new questions. They have been around in one form or another for the past 50 years, as educational researchers and policymakers have tried linking student learning with teaching for the purposes of “increased productivity” or for discerning “school or teacher effects” on student learning (p. 4). Most of these efforts have attempted a precision that would make them acceptable in a psychometric sense, while the inherent messiness of the human contexts under study has continued to ensure varying degrees of failure in obtaining reliability, validity, or perhaps most important, credibility. The aim of this book is to examine a new round of attempts to establish clear linkages between teacher work and student learning by focusing on “school evaluation for accountability and organizational control” (p. 7). And as Millman and Schalock indicate, the jury remains out on whether these recent attempts will be judged as successes or failures. In a sense, the balanced presentation in this neatly edited volume provides an invitation for readers to join in the deliberations.

The book presents four contemporary cases of large-scale accountability plans that attempt to explicate the influences of schools and the work of teachers on the resulting
academic performances of students. These plans are now in place or being considered in Kentucky, Oregon, Tennessee, and Dallas, and they represent a variety of assessment data and methods of analysis. In Tennessee and Dallas, where the focus is on ascertaining school and teacher effects on learning, two different statistical approaches are used to analyze student academic gains on criterion-referenced tests. While the Dallas approach attempts to incorporate multiple assessment devices, Tennessee’s approach provides a longitudinal view of value-added learning drawn from a single nonredundant instrument applied at regular intervals. In Oregon and Kentucky, where much of the data are derived from student performance assessments, evaluation is of teacher effect and school effect, respectively. Each of the plans is provided a section of the book that consists of five chapters. The introductory chapters for each scheme are written by stakeholders who offer overviews from within the political contexts that the plan evolved. This is followed by an explanation of the plan itself, which attempts to present the intent, the methodology, and the results that have accrued. Next come two critiques from outside reviewers whose training and experience provide a basis for acknowledging the possibilities and limitations of the approach. The critiques are generally more civil than provocative—evidence, perhaps, that the reviewers are focusing less on ideological issues than on methodological ones. Each section concludes with a rejoinder by the presenters. What results is a set of balanced overviews and informed reactions that provide an effective digest of information on current large-scale and high-stakes assessment practices.

Fortunately for readers like myself who are more interested than competent in appreciating the statistical and numerical manipulations involving the four schemes, the concluding section of the book provides authoritative summary statements, cautionary notes, and future directions that are intelligible even to the quantitative neophyte. Jason Millman, Linda Darling-Hammond, and James Popham provide balanced and insightful commentaries that represent different shades along the spectrum of assessment philosophies. As such, this last section, marked by its reserve, insight, and humor, deserves attention as a place to begin as well as a place to conclude.

As to the understandability of the four methodologies, it is hard to say if more space would have helped in explaining the various techniques described here. For those curious enough to find out, they may refer to the broader discussions that are cited in the references of each of the four presentations. What is clear is that all four approaches depend to some degree on “statistical gyrations,” as Popham refers to them, to make their cases. What is made clear, too, is that these manipulations sometimes render the schemes indecipherable to their intended constituencies or introduce unintended results that threaten the positive effects that the accountability plans were meant to bring about.

The intellectual progression of the field of educational measurement and evaluation may be gauged by the continuing refinement of methods to broaden the repertoire of learning tasks that can be accounted for validly, reliably, and fairly. While each of the methods outlined and critiqued here displays its own set of limitations and flaws, all of them contribute in one or more ways toward clarifying or expanding the boundaries of evaluation practices.

The Tennessee system and, to some extent, the Dallas system attempt to shift the focus from raw test scores to the added value of taking teacher and school contexts into account