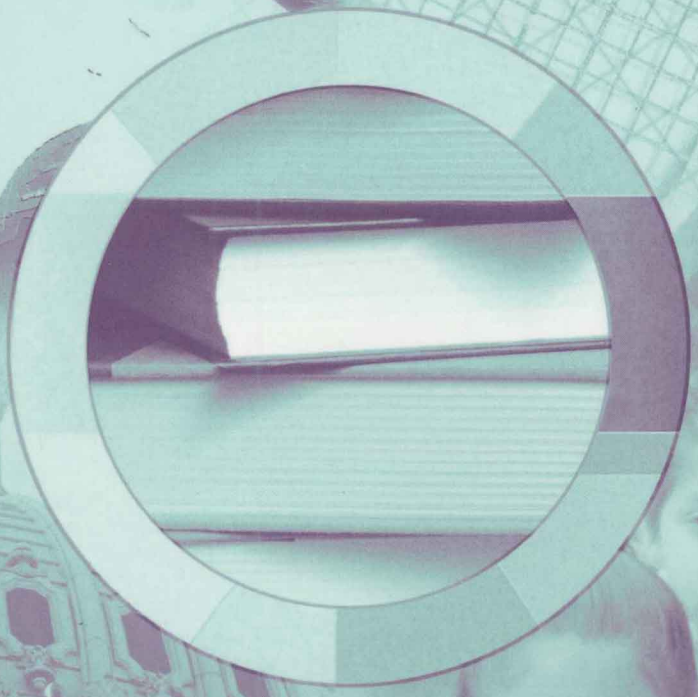
ASSESSMENT IN EDUCATIONAL REFORM

Both Means and Ends



ROBERT W. LISSITZ WILLIAM D. SCHAFER

Assessment in Educational Reform

Both Means and Ends

Robert W. Lissitz

University of Maryland College Park

William D. Schafer

Emeritus, University of Maryland College Park

Allyn and Bacon

Boston London Toronto Sydney Tokyo Singapore

Series Editor: *Arnis E. Burvikovs*
Editorial Assistant: *Matthew Forster*
Production Administrator: *Gordon Laws*
Composition and Prepress Buyer: *Linda Cox*
Electronic Composition: *Peggy Cabot, Cabot Computer Services*
Manufacturing Buyer: *Julie McNeill*
Cover Administrator: *Kristina Mose-Libon*
Marketing Manager: *Kathleen Morgan*

A Pearson Education Company
75 Arlington Street
Boston, MA 02116

Internet: www.ablongman.com

Between the time Website information is gathered and then published, it is not unusual for some sites to have closed. Also, the transcription of URLs can result in unintended typographical errors. The publisher would appreciate notification where these occur so that they may be corrected in subsequent editions.

Library of Congress Cataloging-in-Publication Data

Assessment in educational reform : both means and ends / [edited by] Robert W. Lissitz, William D. Schafer.
p. cm.
Includes bibliographical references and index.
ISBN 0-205-33269-2 (pbk.)
1. Educational tests and measurements—United States—Congresses.
2. Educational change—United States—Congresses. I. Lissitz, Robert W.
II. Schafer, William D.

LB3051 .A769 2002
371.26'0973—dc21 2001022762

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1 06 05 04 03 02 01

Assessment in Educational Reform

prologue

In the spring of 1999, I took a sabbatical in Greece and while visiting Athens attended a conference in honor of a faculty member retiring from a statistics department. There, as I was enjoying the conference, I thought about what an honor this was for the person who was retiring. He must have done some fine things at the University of Athens to motivate people to create a conference in his honor and to travel such distances to show their respect. That was when I realized that such an activity would be the perfect honor for Dr. William Schafer, who had just decided to retire from the University of Maryland. This book is the result of the decision to organize a *festschrift* for Bill.

The topic of the conference and the resulting book is assessment. While on leave from the University of Maryland, Bill served as the director of the State Assessment Office of the Maryland State Department of Education. In that capacity, Bill faced a number of challenges and met them all with success, but in conversations with me, we both realized that there were a large number of unresolved issues and unobtained goals. State assessment is one of the current hot buttons in America. Nearly every newspaper has an article about testing, which includes such topics as accusations of bias, allegations that too much time is allocated to testing, complaints that the right topics are not covered on the test, and questions about whether the tests are used for the right purposes, and so on. The testing endeavor attracts the attention and concerns of politicans, parents, students, and the whole professional education establishment, as well as those who report and comment on such matters. It is not hard to see why assessment is a hot button issue. It is an indicator of success (and failure) of the student and, by implication, the instructional system in which that child should be thriving. These indicators need to be fair and valid to justify so much faith in them. And it takes a lot of faith to use test results to certify competence for graduation or to allocate salary raises, just two examples of important decisions that some people suggest should be informed by test data.

When I returned to the University of Maryland from my sabbatical, I asked Bill Schafer if he would help me put together a conference, at which he would be one of those to address the audience, and then work with me to produce this book. Asking someone to help in a process that honors himself is a little awkward, but we focused on the chance to contribute to the profession to which Bill has devoted his life. Bill's continuing investment in the assessment field made the choice easy—he agreed. This book emanates from that agreement in the fall of 1999, the conference that occurred in the summer of 2000, and the addition of a series of four original papers written for this book.

The conference and the resulting book started as a festschrift but soon became an effort to influence what people actually do in the field as they work to improve the assessment enterprise. We are in a profession that is committed to facing problems and solving them, using formal approaches that are based in sound theory and carefully obtained data. This effort asked the contributors to base their chapters on sound data and the literature that exists in our field, but we asked them to do something else, too. We asked that they speculate and that they try to formulate a sense of best practice. Each of the contributors

has many years of experience in the field and they have a lot to offer. This book tried to give them a vehicle for expression.

Each of the seven chapters that were part of the original conference (those by Drs. Beaton, Brookhart, Hambleton, Roeber, Schafer, Stiggins, and Wise) and the four additional chapters (those by Drs. Airasian, Diez, Stansfield, and Tindal) that were invited to be a part of the book are focused on answering an important question about assessment practice. Bill and I chose the questions and some of the contributors suggested modifications, with the intent of providing an analysis of an issue in applied assessment that would result in a set of observations and recommendations that would help a broad range of education practitioners. We hope the reader agrees that the contributors have succeeded admirably.

Robert W. Lissitz
University of Maryland

preface

The concept of this book, and its value, rests on two assumptions. One is that the potential of assessments in educational practice is very much unmet. The other is that school reform will not be effective until assessments become the focus of reform efforts. The book's purpose is to help direct that focus.

Each author was asked to postulate a new world of excellence in education and to describe what assessments will look like in that world. That is an unusual assignment for people like us. We are conditioned by our profession to be careful, making no claims that cannot be amply justified and even then expressing ourselves in language that makes it clear we understand we might be wrong. The only things we can say are true are those we ourselves see. But here, professionals steeped in that tradition were asked to speculate and describe what none of us has seen. Fortunately, all took that on as an intriguing assignment. As editors, Bob Lissitz and I marveled at their creations.

Assessment exists at numerous levels and with various scopes. Even describing all the ways in which assessments are used educationally would be a daunting task. In order to organize our thinking about topics to include here, we thought of assessments vertically, from those used instructionally by individual teachers with individual students to those used to make national and international comparisons of educational quality by policy makers. Even that is a broad range.

At the level where learning actually takes place, a teacher interacts with a student and each has assessable states and expectations during the instructional process. Susan M. Brookhart and Richard J. Stiggins explore those interactions from the perspectives of the teacher and the student, respectively. To round out building-level uses, Peter W. Airasian and Lisa M. Abrams consider the roles of assessments in schools. Mary E. Diez considers how assessment will be used in evaluating teacher candidates as well as practicing teachers.

State-level assessments are becoming more and more important. My contribution describes how state-level, large-scale assessments might be used to foster school reform. Of particular concern are problems of inclusion of all students in assessment. Gerald Tindal considers this problem from the perspective of students with disabilities, and Charles W. Stansfield and Charlene Rivera discuss approaches to the assessment of students who are non-native-English speakers.

Lauress L. Wise and R. Gene Hoffman discuss how assessment will be used to document the effects of educational reform efforts, and Edward D. Roeber considers the needs of policy makers for assessment information at all levels. Ronald K. Hambleton takes up how the data are presented to enhance interpretation, and Albert E. Beaton discusses overcoming the special problems of interstate and international comparisons.

There were many individuals and institutions contributing to the success of our work. Clearly among those are the College of Education at the University of Maryland, College Park under the leadership of Dean Edna Szymanski and the Department of Measurement, Statistics, and Evaluation under the leadership of my co-editor, Robert W. Lissitz. Also invaluable was the support of the Maryland State Department of Education, and especially

Nancy S. Grasmick, the State Superintendent of Schools, and Assistant State Superintendent Mark Moody. We also are indebted to our publisher, Allyn & Bacon, for its willingness to take this rather futuristic project on as well as for its careful and helpful work along the way.

I hope you agree that this thought-provoking collection contains more than just lip service to meaningful educational reform. And I hope you enjoy it as much as I did helping to put it together!

William D. Schafer
Emeritus, University of Maryland

contents

Assessment in Educational Reform

chapter

1 What Will Teachers Know about Assessment, and How Will That Improve Instruction?

SUSAN M. BROOKHART

Duquesne University

This book is dedicated to visualizing ways in which assessments can be instrumental in effecting positive change in educational institutions. I have been asked to focus on teachers: In an ideal (or at least improving) world, what would teachers know about assessment, and how would they use that to improve instruction? Let me air my bias right from the outset. I think I was assigned one of the most important questions, because the education we all wish to improve happens in classrooms, by virtue of the work teachers and students do. Administrators make it possible for education to happen, and indeed quality classroom learning can't happen without the resources and policies they administer. But the locus of the teaching–learning experience is in the classroom, and its primary agents are teachers and students. It follows that what teachers and students know and do about assessment is critical.

The most important assessment information for classroom use comes from classroom assessments (see Stiggins, this volume). In this chapter, I have been asked to address teachers' knowledge and use of large-scale assessment information. I acknowledge that for teachers, large-scale assessment information is—and should be—secondary to classroom assessment information. Nevertheless, I can envision some important ways in which teachers' knowledge and use of large-scale assessment information would improve instruction. Some of these ways involve changing teachers' knowledge or practices, and some of them involve changing the kind of large-scale assessment information we routinely offer to teachers so that it is more useful for improving instruction.

Framework for Posing the Question: What Will Teachers Know and How Will That Improve Instruction?

Before I address this question, I need to deal with the underlying assumptions. If I just answered it, I would be participating in what Cochran-Smith and Lytle (1999) called "knowledge-*for*-practice"; that is, I would be disseminating research- or expert-based theoretical knowledge for teachers to apply to the examples and instances their classroom practice presents.

Another approach, "knowledge-*in*-practice," would eschew this whole book. Knowledge-*in*-practice elevates practical wisdom at the expense of formal theory. This approach likewise designates "experts" and "novices"—it just chooses them from a different pool (Cochran-Smith & Lytle, 1999). Teachers become the experts in a domain no one else is expected to understand. Although I understand this approach in the existential sense, it does not lead to stable, shareable knowledge. That, of course, is part of the point of this approach; it sees practical knowledge as continually constructed and reconstructed in action. Knowledge exists in and through the action of teaching, making chapters such as this one moot.

A more productive way to address this question might be to take the approach that Cochran-Smith and Lytle (1999) called "knowledge-*of*-practice." Knowledge-*of*-practice relies on teachers' communal inquiry and is very much about process, politics, and power. The first step is not up to the authors of this book; a group of teachers would decide they wanted to see how understanding and using assessment information would improve

instruction in some specific way, but they would be free to use information from many sources. The point would be to see what they could learn. This learning could be shared. It would be judged not according to the canons of academic scholarship, but by its effects on improving learning for all students and improving the practice of teachers. Thus, this approach to teacher knowledge has a democratizing aspect to it.

The Knowledge-*for*-Practice Approach: What We Know

Rather than dismiss the knowledge-*for*-practice approach out of hand, I want to describe briefly what has already been written on the topic of what experts have to say about what teachers should know and be able to do. There are two purposes for this section of the chapter. First, there is actually quite a body of literature under the general theme of investigating "teacher competence" in educational measurement as defined by measurement experts. We have learned some things about teachers' knowledge and use of assessments over the past 40 years, but in this chapter, I argue that the documentation of experts' perspectives on teachers has not succeeded in improving assessment practices and instruction. I think that's because the one-way flow from expert knowledge to teacher application is not the best approach to teacher learning. Cochran-Smith and Lytle's framework helps us see that. But I would like to report selectively on the literature that effectively defines the history of the question this chapter addresses.

Second, one of the purposes of this book is to honor the career of Bill Schafer, who has long been interested in what teachers need to know about assessments in their practice. In fact, one of the first conversations I remember having with Bill was about an assessment workshop he was doing with teachers. Citing the "teacher competence in assessment" literature affords me the opportunity to cite Bill's work. Much to his credit, the Schafer articles have titles like "Assessment Literacy for Teachers" (Schafer, 1993, after Stiggins, 1991a, who also called it "Assessment Literacy") and "Essential Assessment Skills in Professional Education for Teachers" (Schafer, 1991)—much more gracious language than "teacher competence."

There are several good reviews of the literature concerning what teachers know about assessment. In 1989, the Buros–Nebraska Symposium on Measurement and Testing addressed the question, "Are our school teachers adequately trained in measurement and assessment skills?" An edited volume of papers on this topic (Wise, 1993) comprises a comprehensive review of literature about what preservice teachers are taught about measurement, what they should be taught, what they actually know, and the observed quality of their assessment practices. Two chapters are particularly relevant for our purposes: Gullickson's (1993) review of literature on teachers' attitudes and practices and the content of undergraduate measurement courses, and Marso and Pigge's (1993) review of literature on teachers' attitudes and practices and studies of direct assessments of teachers' testing knowledge. Add to these the literature review in Schafer and Lissitz (1987) about what school personnel should know, do know, and have the opportunity to learn, and an entire special issue of *Educational Measurement: Issues and Practice* (Nitko, 1991) on the topic, and we have plenty of sources for a review of this teacher-competence approach, plus the suggestion that its popularity peaked about a decade ago.

Gullickson (1993) pointed out that measurement experts have weighed in on what teachers need to know about assessment since at least the first volume of the *Journal of Educational Measurement* (Mayo, 1964). Studies of what teachers are or should be taught and what they know have generally concluded that while teachers' knowledge of large-scale testing is limited—especially in the important area of communicating assessment results, arguably something teachers should be able to do (AFT, AERA, & NEA, 1990)—their skills at gathering and using classroom assessment information are much more important. The conclusion of many authors is that measurement courses for teachers should increase emphasis on classroom assessment, and decrease emphasis on large-scale testing (Gullickson, 1993; Marso & Pigge, 1993; Stiggins, 1991b). I wholeheartedly endorse this position. Because of the topic of this chapter, however, I'd like to revisit what some of these studies have concluded about teachers' attitudes, knowledge, and needs regarding large-scale testing.

Teacher Attitudes toward Large-Scale Assessment. Schafer and Lissitz (1987) reviewed teacher surveys of the 1970s and 1980s and concluded that teachers had in general reported a positive attitude toward standardized testing. They noted that this was a well-kept secret and that popular opinion at the time assumed that teachers disliked testing (Schafer & Lissitz, 1987, p. 59). Marso and Pigge (1993) reviewed a large number of studies and concluded there were conflicting findings about teacher attitudes toward assessment. Some studies they reviewed found teachers expressed confidence in their knowledge about testing; some found an expressed lack of confidence. Many of the studies in the review were about "testing" generally. The main focus of Marso and Pigge's own work was classroom testing.

Wise, Lukin, and Roos (1991), using data collected in Nebraska in 1989, found that 77% of teachers rated themselves as "good" or "very good" at interpreting standardized test results; 82% rated themselves as "good" or "very good" at explaining standardized test scores; and 73% believed that possessing strong skills in testing and measurement was important for being perceived as professional. Impara and his colleagues (Impara, Divine, Bruce, Liverman, & Gay, 1991) found some evidence that both measurement instruction and the provision of interpretive information along with results affect teachers' actual abilities at interpreting test results.

Some studies that have taken a different approach than the one I have characterized as the teacher-competence approach have uncovered some negative teacher attitudes toward large-scale testing. Of course, what one finds depends in part on what one looks for. The teacher competency in assessment literature complements a view of large-scale testing that is understood using the language of "competency" or "minimum competency" for students. Many states have begun to talk about large-scale assessment using language about testing for "standards" or "accountability." Researchers' concerns have been about the effects of these testing programs on curriculum, instruction, and teachers.

Studies looking for these effects have uncovered more negativity in teachers' attitudes, but not about the tests themselves so much as their effects on instruction, because that is how the questions have been asked. Smith and Rottenberg (1991) found that consequences of externally mandated testing in elementary schools included a narrowing of curriculum and an increase in instructional time geared to the content and format of the tests. More relevant for this review of teacher attitudes were their findings that teachers

disliked the tests, believing that they caused undue stress and fatigue for their students. Teachers' own emotional responses to tests were reported as shame and embarrassment at low scores but merely relief at high scores, based on a belief that tests measured socioeconomic status as much as achievement (Smith & Rottenberg, 1991). Herman and Golan (1993), also looking at elementary schools, found similarly that testing had effects on curriculum and instruction and that these effects were more pronounced for low-SES schools. Teacher attitudes in this study were somewhat positive in that teachers reported taking responsibility for their students' learning and having some control over it. About the efficacy of testing itself, teachers were not positive, reporting disagreement with statements about testing helping with school improvement, giving useful feedback, or focusing learning goals. Rather, they reported testing caused stress for both teachers and students. Teachers in schools where test scores were improving reported more pressure from the community to raise test scores than did teachers in schools where test scores were dropping—one of the few differences found between those groups (Herman & Golan, 1993).

For our purposes, then, we can conclude that teachers have mixed attitudes toward the enterprise of testing itself, including some positive attitudes about test information that is used well. They have negative attitudes toward tests used in such a way as to have what they perceive as negative consequences for their students. They do not perceive their own knowledge about testing to be a major problem.

Instruction about Large-Scale Assessment. Let's turn now to what others think teachers ought to be taught about large-scale assessment. I summarize this literature under the teacher-competency umbrella because this literature is about what various expert groups think teachers ought to know.

There are several approaches to the question of what teachers, either inservice or preservice, should be taught about assessment, and, taken together, they tell us quite a lot. Methods and approaches have included surveys of teachers and of measurement experts, reviews of program and course requirements, reviews of textbook contents, and observation and job analysis of teaching. As with teacher attitudes, the literature on content of instruction has been well reviewed, by some of the authors in this book and by others. Generally, surveys of teachers and the job-analysis approach have concluded by recommending more instruction, at both the preservice and inservice level, to build a repertoire of strategies for high-quality classroom assessment, and less instruction in standardized testing than is currently the case in most measurement courses for teachers. Textbook reviews or surveys of measurement experts have placed more emphasis on learning about standardized testing, although some measurement experts who have studied the classroom and its information needs do not (Airasian, 1991). Recommendations for instruction in standardized testing often include a shift away from statistical interpretation of scores and toward communication of score meaning to a lay audience (Brookhart, 1999; Schafer, 1991; Schafer & Lissitz, 1987).

Gullickson (1993) summarized the results of several survey studies he and his colleagues conducted in the early 1980s. In his summary, he compared the desired content emphases of teachers and professors by lining up the 20 content priorities rated most highly by each group. There was a match in only one of five areas, topics dealing with preparation of classroom tests. Professors reported more coverage of test statistics and

analysis and of standardized test applications than teachers desired, and teachers desired more topics covered under formative and summative uses of tests and nontest evaluation practices than professors reported (Gullickson, 1993, pp. 12–13).

Are the data dated? After all, today's undergraduate teacher education students were infants in the early 1980s. In his editorial introduction to a special issue of *Educational Measurement: Issues and Practice*, Smith (1999) suggested that the lack of relevant measurement instruction has remained a problem. He wrote, "A lack of interest in the issues that concern practicing educators on the part of the measurement community puts us in danger of ceding any influence in what classroom assessments look like and how they are used to our subject-matter colleagues" (Smith, 1999, p. 4). This is part of a strongly worded argument in favor of more emphasis on classroom assessment and less on large-scale assessment in measurement courses for teachers. And even though I argued, in the same issue, for a small place in teacher education for using and communicating the results of large-scale assessments (Brookhart, 1999), I agree completely that the emphasis in such courses should be placed on classroom assessment.

The *Standards for Teacher Competence in Educational Assessment of Students* (AFT, NCME, NEA, 1990), arguably the normative reference for recommended content in assessment courses for teachers, includes seven standards. Teachers should be skilled in:

1. choosing assessment methods appropriate for instructional decisions;
2. developing assessment methods appropriate for instructional decisions;
3. administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods;
4. using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement;
5. developing valid pupil grading procedures which use pupil assessments;
6. communicating assessment results to students, parents, other lay audiences, and other educators; and
7. recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information (AFT, NCME, NEA, 1990).

Standards 3, 4, 6, and 7 include skills that apply to large-scale assessment, including administering, interpreting, and communicating assessment results, using information for decision making, and recognizing unethical practices. The recommendations of Schafer (1991) and Stiggins (1991b) both parallel these content areas; although they differ in the amount of emphasis they put on teachers developing skills related to large-scale assessment, they both acknowledge teachers' need for these skills.

Knowledge of Large-Scale Assessment. Some studies have actually investigated teachers' knowledge about assessment. Two methods have been used: tests of assessment knowledge or reviews of teachers' assessments themselves. The latter studies have led to conclusions about teachers' need for instruction in classroom assessment (Marso & Pigge, 1993). More relevant for our review are the studies in which teacher knowledge was tested, because these have included questions about standardized tests as well as classroom assessments.

Gullickson (1993) pointed out that one of the strongest arguments for teachers knowing about standardized tests is simply that they are given in almost all schools. Teachers