

1

Overview of the Book

Thomas A. Romberg

The purpose of this book is to share with mathematics educators a set of recent papers written on issues surrounding mathematics tests and their influence on school mathematics. The impetus for the contributions grew from a conference, "The Influence of Testing on Mathematics Education," sponsored by the Mathematical Sciences Education Board (MSEB) at UCLA in June 1986. The purpose of the conference was to gather informed input and advice on current testing practices. The fact is that students in American schools are subjected to a variety of tests, often standardized tests, from kindergarten to graduate school. Such tests are, according to widely held perception, inhibitors to change and improvement in education and especially in mathematics education. Since MSEB was organized to coordinate the current reform movement in school mathematics, the topic of the conference was deemed critical. Two things became clear at the UCLA meeting: First, there was agreement that tests need to change to reflect curriculum changes, and second, many participants articulated their beliefs about the inadequacy of current tests and provided relevant anecdotes on problems to others at the conference. However, no one was sure how such changes could be accomplished, nor did participants even have substantive, reliable information about the actual impact of testing on classroom practices.

Following the conference, a three-person Testing Design Task Force, consisting of Jeremy Kilpatrick, Tej Pandey, and Thomas Romberg, was organized by MSEB. In September 1986, this task force produced "A Proposal for Studies of Mathematics Tests and Testing" based on the following four assumptions:

- Valid information about student mathematics is needed by a variety of people (students, teachers, parents, administrators, policy makers) for a variety of purposes (monitoring progress, selection for and placement in courses, program evaluation, accountability).
- Both the curriculum and teaching practice in mathematics need to be directed toward strategies which students could use to solve problems, the application of mathematics to practical situations, and the development of thinking skills. Consequently, testing should reflect students' achievement in these directions.
- Serious questions have arisen about the validity of existing tests for the uses to which they are being put. Standardized tests and state-mandated tests may yield information that is invalid for certain purposes and provide little or no information on several important dimensions of achievement.
- The continued use of existing tests appears likely to impede the much-needed reform in curriculum and instruction to which the mathematics education community is committed.

On the basis of these assumptions, a set of questions and research studies was proposed. In particular, several literature reviews were planned, each of which would explore one facet of the validity of mathematics tests for various purposes. Topics were to include surveys of testing practice, the alignment of tests with curricula, test-preparation practices and effects, test-taking skills, the student use of calculators during test taking, teacher and student attitudes toward tests, time spent in testing, alternatives to testing, and minority group and gender group differences in risk taking and test performance.

In 1987, when the Wisconsin Center for Education Research was awarded the grant to form the National Center for Re-

search in Mathematical Sciences Education (NCRMSE), NCRMSE assumed responsibility for carrying out several aspects of the proposed scope of work outlined by the MSEB task force. The papers in this volume represent a number of the literature reviews proposed. They were written by the Center staff or by invited scholars. The contributions cover many of the issues identified by the MSEB task force and are an important contribution to our growing knowledge about the impact of tests and testing on school mathematics.

The papers are only a part of the work now being conducted by the Center on this important topic. Since 1987, the Center has conducted two major surveys. The first, a national survey of a sample of Grade 8 mathematics teachers (Romberg, Zarinnia, & Williams, 1989), provides information about teachers' perceptions of the impact of mandated testing on their instruction. Findings reveal that teachers are familiar with mandated tests, make efforts to ensure that students perform well on the tests, and adjust their curriculum and modes of instruction to focus on the knowledge and skills being tested.

The second is a survey of state mathematics coordinators on the current types of mandated testing in the fifty states (Romberg, Zarinnia, & Williams, 1990). This study examines the actual mandated testing practices in each state, including the kinds of tests given, the uses to which they are put, and the kinds of test-score information subsequently available to the teachers.

In addition to these surveys and this collection of papers, four related activities are now in progress:

- 1) During the past year, two in-depth case studies on the impact of mandated testing in classrooms have been conducted at four sites. Information from these studies is now being analyzed.
- 2) Three extensive reviews of literature and of practice are now underway on classroom testing for instructional decision making, testing for placement and grouping, and test validity.
- 3) Some sample test items have been written and are being tried out; they have been designed to assess level of reasoning in some of the particular domains outlined in the *Curriculum and Evaluation*

Standards for School Mathematics (National Council of Teachers of Mathematics, 1989).

- 4) Two projects on curriculum design, development, and assessment are being conducted jointly with the Research Group in Mathematics Education at the University of Utrecht.

In all, the work of the staff and consultants of NCRMSE on the influence of testing on school mathematics makes it clear that valid information about student performance is sorely needed if the reform movement in school mathematics is to succeed.

OVERVIEW OF THE CHAPTERS

The next twelve chapters were prepared during 1988 and 1989. Chapter 2 represents a summary of the overall problems associated with the need for valid information. Chapters 3 and 4 examine the use of tests in the context of the current reform movement in school mathematics. Chapters 5 and 6 describe the current procedures used to develop state tests. Chapter 7 summarizes current efforts to incorporate the use of calculators in mathematics tests. This is followed by chapter 8, a review of research on testing with calculators. Chapter 9 is a review of gender differences and testing. Chapter 10 is an examination of an Australian project addressing teachers' assessment practices. The next two chapters, 11 and 12, deal with alternative strategies for gathering, analyzing, and reporting student performance information. The final chapter is an invited review and critique of chapters 2 through 12.

Chapter 2: Evaluation: A Coat of Many Colors by Thomas A. Romberg

An earlier draft of this chapter was prepared as an invited address for Theme Group—T4, Evaluation and Assessment, at the Sixth International Congress on Mathematical Education in Hungary. This paper examines both the methods of gathering information from students and the use of that information to make a variety of judgments. It considers the history of evaluation and how evaluation relates to the gathering of assessment data and to educational decision making. To examine the

strengths and weaknesses of the evaluation of the impact of new mathematics programs and of large-scale profile evaluations, it describes trends in evaluation and assessment that show the disparity between what is possible and what is, in fact, achieved.

Chapter 3: Implications of the NCTM *Standards* for Mathematics Assessment

by Norman Webb and Thomas A. Romberg

In 1987 work began on the NCTM *Standards*. Thomas Romberg was the chair of the commission that produced this document and Norman Webb chaired the working group that prepared the evaluation standards. This chapter includes criteria for assessment which would be compatible with and supportive of the curriculum standards. Three examples of alternative assessment techniques are presented that correspond to the intent of the evaluation standards and provide illustrations of forms of assessment that are applicable in evaluating the curriculum standards.

Chapter 4: Curriculum and Test Alignment

by Thomas A. Romberg, Linda Wilson,
'Mamphono Khaketla, and Silvia Chavarria

In this chapter, a variety of tests and test items are examined to determine whether they reflect the recommendations made in the *Standards*. In the initial sections, six commonly used standardized tests are examined. It is clear from this examination that those tests fail to assess the higher-order skills such as problem solving, reasoning, and connections that are stressed in the *Standards*. Then items are identified from other tests which could be used to assess such aspects of mathematics.

Chapter 5: State Assessment Test Development Procedures

by James Braswell

The primary purpose of this paper is to describe how tests are developed for state assessment programs. The methods described are based in part on discussions with state department of education assessment staff members in Florida, Louisiana, Massachusetts, Michigan, and New Jersey—states in which testing

practice was judged to be representative of a range of approaches to test development. Occasional references that reflect previous experience with other state testing programs and current work with the National Assessment of Educational Progress test development team for the 1990 Mathematics Assessment are also taken into consideration.

Chapter 6: Test Development Profile of a State-Mandated Large-Scale Assessment Instrument in Mathematics
by Tej Pandey

Two main types of large-scale assessments are examined in this paper. The first focus of interest is oriented primarily to those individuals who typically use test information to rank a student on an established norm, find a student's strengths and weaknesses, and determine whether that student has mastered specific course content. The second focus of interest lies primarily in the administrative use of information to determine the achievement level of students in a school, district, or regional system for purposes of assessing program effectiveness. This paper examines the nature and design of test instruments in a large-scale assessment program (California Assessment Program) providing reliable group-level information. The paper also describes the test development process as it has evolved over a period of fifteen years to meet the curriculum demands of the time.

Chapter 7: Assessing Students' Learning in Courses Using Graphics Tools: A Preliminary Research Agenda
by Sharon L. Senk

Recently mathematics educators have called for the use of calculator and computer-graphing technology in mathematics classes, and several software and curriculum development projects have been initiated to transform these recommendations into reality. However, until now, there has been little systematic study of how teaching, learning, and assessment in courses using such graphics tools are affected by the technology. This paper describes a preliminary agenda developed by researchers in the field for assessing students' learning in courses using graphing tools. Included are suggested investigations of student and teacher outcomes and a discussion of methodological issues.

**Chapter 8: Mathematics Testing with Calculators:
Ransoming the Hostages**
by John G. Harvey

This paper argues that present testing practices hold today's students hostage to yesterday's mistakes. The author predicts that because mathematics tests fail to incorporate the use of calculators in the testing process, mathematics instruction will fail to incorporate the use of calculators effectively, continuing to hold today's students prisoners to a mathematics curriculum that is failing to prepare them for society's immediate needs as well as those of the twenty-first century. The paper suggests that the use of calculators on mathematics tests will not remedy the failures of present tests, but that their use is necessary if we want students to investigate, to explore, and to discover mathematics effectively.

Chapter 9: Gender Differences in Test Taking: A Review
by Margaret R. Meyer

Ideally, when students take a mathematics examination, the only thing that should influence their score is their mastery of the material being tested. This paper reviews evidence concerning the existence of gender differences in mathematics test taking. It examines several factors that have surfaced relating to differences in performances for males and females. These factors are power vs. speed test conditions, item-difficulty sequencing, examination format, test-wiseness, risk-taking behavior, and test-preparation behaviors. One conclusion reached is that the use of the multiple-choice format could result in a male advantage. A recommendation is therefore made that assessment instruments not rely as heavily on the multiple-choice format.

**Chapter 10: Communication and the Learning of
Mathematics**
by David Clarke, Max Stephens, and
Andrew Waywood

The learning of mathematics is fundamentally a matter of constructing mathematical meaning. The environment of the mathematics classroom provides experiences which stimulate this process of construction. This chapter presents the findings of

three studies based in Australian schools: the IMPACT Project, Assessment Alternatives in Mathematics, and the Vaucluse College Study. The purpose of the research synthesis considered in this chapter is to discuss (a) the extent to which the strategies reported encourage children to broaden their mathematical thinking and facilitate meta-learning and (b) the impact of these strategies on the nature of mathematical activity in classrooms, in particular with reference to redefining the roles of teacher and student in creating and giving personal meaning to mathematics.

Chapter 11: Measuring Levels of Mathematical Understanding by Mark Wilson

This chapter describes recent psychometric advances in the creation of models that measure developmental change in understanding. Standardized, norm-referenced tests are based on an accumulation of bits of knowledge rather than on understanding, which is a constructivist, developmental process. As the latter conception gains more acceptance, there is a need for new assessment models. Empirical examples of response maps are used to illustrate the potential of the new models.

Chapter 12: A Framework for the California Assessment Program to Report Students' Achievement in Mathematics by E. Anne Zarinnia and Thomas A. Romberg

The purpose of this paper is to propose categories for the California Assessment Program that report student achievement in mathematics. Initially, the purpose of reporting achievement was accountability. This paper examines explicit and tacit messages imposed in the analyzing, gathering, and aggregating of this information that expose subtle effects on teaching and student achievement. The paper determines that units of analysis and reporting categories are needed that will both deliberately support the purposes of gathering adequate information for monitoring and—by focusing attention on critical considerations—promote reform in mathematics education. This chapter outlines seven bases for forming reporting categories.

Chapter 13: Evaluation—Some Other Perspectives by Philip C. Clarkson

A common response to the challenge of the *Standards* is, "Yes, but who will change the tests?" (National Council of Teachers of Mathematics, 1989, p. 189).

It is apparent that "the tests" referred to are not the tests teachers give in their classrooms on a day-to-day or weekly basis. They have control over those already. "The tests" are the standardized assessment instruments which are used throughout the United States, often authorized by legislation, devised by commercial organizations, and seen by many teachers in the country as being a forceful factor in structuring their mathematics curriculum.

The preceding chapters have provided background information on these tests and have made some suggestions on how they could be altered. None reflect on the question as to whether they are indeed necessary. This chapter sketches developments over the last twenty-five years in the State of Victoria, Australia, where there is now only one external test given at the end of the school system, in Year 12. This contrasting situation may contribute constructively to the ongoing debate in both Australia and in the United States as to how to monitor the work of schools.

In summary, as the title to this book suggests, the authors of these chapters address an important set of issues about mathematics assessment and evaluation. It is clear that it is important to gather information on student performance in mathematics for a variety of reasons. However, while the mathematics curriculum and the way mathematics is taught are changing, the definition of assessment and how performance is assessed also need to change. It is imperative, if the school mathematics reform efforts are to be successful, that mathematics educators become aware of the issues addressed in these chapters.