

Testing for Learning

HOW NEW APPROACHES
TO EVALUATION CAN IMPROVE
AMERICAN SCHOOLS

Ruth Mitchell



Testing for Learning

HOW NEW APPROACHES
TO EVALUATION
CAN IMPROVE
AMERICAN SCHOOLS

Ruth Mitchell



THE FREE PRESS
A Division of Macmillan, Inc.
NEW YORK

Maxwell Macmillan Canada
TORONTO

Maxwell Macmillan International
NEW YORK OXFORD SINGAPORE SYDNEY

Copyright © 1992 by Ruth Mitchell

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the Publisher.

The Free Press

A Division of Macmillan, Inc.
866 Third Avenue, New York, N.Y. 10022

Maxwell Macmillan Canada, Inc.
1200 Eglinton Avenue East
Suite 200
Don Mills, Ontario M3C 3N1

Macmillan, Inc. is part of the Maxwell Communication
Group of Companies.

Printed in the United States of America

printing number

1 2 3 4 5 6 7 8 9 10

Multiple-choice tests reprinted with the permission of the Psychological Corporation and The Riverside Publishing Company.

Enhanced multiple-choice illustration, writing prompt, standards for scoring writing and open-ended mathematics, and history examination reprinted with the permission of the California Assessment Program, California State Department of Education.

Humanities exhibition reprinted with permission from the Center for Collaborative Education, New York.

Purposes for portfolio assessment reprinted with permission from Portfolio Assessment Clearinghouse, San Dieguito Union High School District, Encinitas, California.

Questionnaires for students and parents reprinted with permission from Kathryn Howard, Pittsburgh Public Schools.

Criteria for mathematics portfolio contents reprinted with permission from the Vermont State Education Department.

The Primary Language Record reprinted with kind permission from the Centre for Language in Primary Education, Webber Row, London, England.

Library of Congress Cataloging-in-Publication Data

Mitchell, Ruth

Testing for learning : how new approaches to evaluation can improve American schools / Ruth Mitchell.

p. cm.

Includes bibliographical references (p.) and index.

ISBN 0-02-921465-3

1. Educational tests and measurements—United States.

2. Education—United States—Evaluation. I. Title.

LB3051.M634 1992

371.2'6—dc20

91-27896

CIP

*To the teachers,
who know and love their students*

Preface

Evaluation sends a message. It points to what is valued and ignores what is not perceived to be important. Educational evaluation—testing and assessment—has been telling students, teachers, administrators, and legislators that the system values rote memorization and passive recognition of single correct answers. This message has been powerfully conveyed by the ubiquitous multiple-choice tests which have dominated American educational evaluation for most of the past thirty years and have terrorized it in the 1980s.

Worse than the form of the tests themselves has been the message that a single test can determine what students know and can do. Multiple-choice tests would not be so bad if they were part of a spectrum of evaluations, including essays, cooperative productions, collections of work, and teachers' observations. But evaluation has narrowed to the "bubble" on a machine-scorable answer sheet.

At the same time as testing has been distorting what is taught and learned, turning it into pellets which are the intellectual equivalent of rabbit food, other forces have been pressuring schools to move in the opposite direction. Business executives and leaders of industry find that they want employees who can think for themselves and apply knowledge to new situations. Professional associations of teachers, such as the National Council of Teachers of Mathematics, the National Council of Teachers of English, the National Science Teachers' Association, and the National Council on the Social Studies (to name only a few), have rethought what they were teaching as part of their professional responsibility. These associations and others like them have all researched, written, and

published curriculum frameworks within the past five years. The frameworks share a common emphasis on thinking, problem solving, conceptual understanding, solid academic knowledge, and the application of learning.

These components of a sound academic education have been advocated for thirty-five years by the Council for Basic Education, which enthusiastically endorsed the new curriculum frameworks, especially the National Council of Teachers of Mathematics' *Curriculum and Evaluation Standards for School Mathematics*. The Council's president, A. Graham Down, perceived a collision between what should be taught in schools and how it is tested. With his customary vision, he sought ways to promote evaluation compatible with the rich intellectual experience schools should provide. He asked me to look at the newly emerging kinds of evaluation which ask students to demonstrate directly, not through the proxy of a "bubble," what they know and can do and then to write a book about them.

As an advocacy organization, the Council operates largely through the written word, publishing studies of educational developments, a monthly journal of comment, and occasional papers examining specific educational topics. As befits the president of an organization that has lived on "soft" money for thirty-five years, Graham also sought financial support for the project, and received it from the National Science Foundation; The Lilly Endowment, Indianapolis; and the McKenna Foundation, whose generosity is gratefully acknowledged.

Assessment is as old as education itself. For most of its history, educational assessments consisted of recitations, oral demonstrations of mastery over a subject, or essays. In the United States assessment took these forms (and still does in some, mostly private, schools), until the middle of this century, when multiple-choice took over. By the 1980s, the bubble had eclipsed nearly all other forms of assessment in U.S. public schools. How students are assessed inevitably affects how they are taught. Assessment cannot be considered separately from teaching and learning, because assessments are the motivation for both teacher and student. Changing assessment therefore impacts on the classroom, the textbooks, the professional lives of teachers, the decisions of administrators. The topic here is a new system, not just modified tests.

I wrote the book for a general audience, not primarily for educational specialists in university departments of education, or in school

or state education administrations, but for teachers, parents, school board members, taxpayers, legislators and their staffs, journalists, professional education-watchers, but above all the large number of general readers who care about what is happening in U.S. schools as a vital component of the economic, social, and political fabric of our civilization. The book largely consists of descriptions of programs, and weaves into them discussions of theoretical issues. Not all issues are treated in each case, but only where the circumstances presented an opportunity.

After a brief introduction to the players and the action on the evaluation scene, the book plunges straight into examples. It concludes with a brief historical description of how we got ensnared by a single form of testing, and some thoughts on how a broad-minded system can move our schools toward their goals.

I have consciously tried to use general language, not the technical jargon of education. The word “curriculum” appears only rarely in this book, when it absolutely cannot be avoided or appears in a quotation. “Teaching and learning” are used instead in order to gain both directness and accuracy. Philip Schlechty, president of the Center for Leadership in School Reform in Louisville, says in his 1989 book, *Schools for the Twenty-First Century*, that educational evaluation is “too important to turn over to the measurement specialist.” I hope to persuade the general reader that educational assessment is not an arcane subject, but everyone’s business.

Many people have contributed to this book, as I travelled around the United States talking to educators, attending conferences, watching students as they performed tasks. I want to thank everyone on the following list for their patience with my requests for information. The list is as complete as memory and records can make it; if anyone’s contribution has gone unrecognized, please forgive me.

Walter Askin, California State University, Los Angeles

Joseph Flynn, Cleveland Education Foundation

Joan Lipsitz, The Lilly Endowment

David Florio, formerly of the National Science Foundation, and now of the Office of Educational Research and Improvement, U.S. Department of Education

Ramsay Selden, Council of Chief State School Officers

Maryellen Harmon, Center for the Study of Testing, Evaluation, and Educational Policy

George Madaus, Center for the Study of Testing, Evaluation, and Educational Policy
Ray Campeau, Bozeman, Montana
Alice Sims-Gunzenhauser, Educational Testing Service (ETS)
Walter MacDonald, Advanced Placement Program, ETS
Deborah Meier, Central Park East Schools
Patricia Bolanos, The Key School, Indianapolis
Grant Wiggins, CLASS, Rochester
Douglas Reynolds, New York State Education Department
Rodney Doran, University of Buffalo
Mary Ann Smith, University of California, Berkeley
Roberta Camp, ETS
Dennis Palmer Wolf, Project Zero, Harvard
Dale Carlson, California Assessment Program (CAP)
Susan Bennett, CAP
The staff of CAP
Richard Mills, Vermont Education Department
Ross Brewer, Vermont Education Department
Geoffrey Hewitt, Vermont Education Department
Robert Kenney, Vermont Education Department
Joan B. Baron, Connecticut Education Department
Stephen Leinwand, Connecticut Education Department
Michael Fischer, New York State Education Department
Joanne Lenke, Psychological Corporation
Fredrick L. Finch, The Riverside Publishing Company
Elaine Craig, Center for Civic Education
Sophie Sa, The Panasonic Foundation
Michael Holtzman, Consultant to the Panasonic Foundation
Gerald Kulm, American Association for the Advancement of Science
Eva Baker, Center for Research in Evaluation, Students, and School Testing (CRESST)
Jan Camplin, Lake Shore School District, New York
Sharon Partyka, Buffalo, New York
Marilyn Whirry, Manhattan Beach, California
Wynne Harlen, Edinburgh, Scotland

Anne Qualter, Liverpool, England
 Myra Barrs, London, England
 Anne Thomas, London, England
 Betty Hagestadt, London, England
 Heather Lewis, Center for Collaborative Education, Central Park East Schools, New York
 Teachers and students of Central Park East Schools, New York
 Michael Goldman, Coalition of Essential Schools
 Kati Haycock, Children's Defense Fund
 Steven Seidel, Project Zero, Harvard
 Renika Zessoules, Project Zero, Harvard
 Lois Easton, Arizona Education Department
 Reuben Carriedo, San Diego City Schools
 Grant Behnke, San Diego City Schools
 Teachers in Cleveland, Ohio, Elementary schools
 Kathryn Howard, Reizenstein Middle School, Pittsburgh PA
 Willa Spicer, South Brunswick NJ Public Schools
 David Sink, University of Birmingham, Alabama
 Janica Loomis, Center for Law and Education, Birmingham, Alabama
 Kathleen Fulton, Office of Technology Assessment, U.S. Congress

I owe a special debt to A. Graham Down for his encouragement and support, and for the same (and for indulgence when I was not in the office, but off enjoying myself watching real children in real schools) to my colleagues: Elsa Little, Barbara Manzon, Patte Barth, Amy Stempel, Catherine Meikle, Maryanne Annan, Stephanie Soper, Michelle Taunton, Karen Anderson, and Liza Benson. Amy Stempel acted as more than research assistant: she was an astute critic and framer of penetrating questions. Patte Barth, Graham Down, and Kati Haycock of the Children's Defense Fund read early versions of the manuscript, and thanks to their perceptive comments, would now find it unrecognizable.

For the first three months of research for this book, I was assistant director of Center for Academic Interinstitutional Programs (CAIP) at UCLA, and was therefore supported by the University of California. My colleagues at CAIP, Patricia S. Taylor, director; Rae Jeane Williams, UCLA Writing Project Director; Susie W. Hakansson,

UCLA Mathematics Project Director; and Janet M. Thornber, UCLA Science Project director, supported me from the beginning and have continued to encourage me since I moved to Washington. Thank you.

Another colleague at UCLA, Mike Rose, has encouraged me and argued with me for more than thirteen years as we shared a common dedication, first to improving writing instruction, and then, as our horizons opened out, to ensuring the best education for all students. Mike showed me how to interest the publishing world in this book, and therefore led to me the person whose patient editing has given it whatever quality it has.

Susan Milmo, my editor at The Free Press, obviously understands how performance assessment works as a feedback mechanism, for she gave me the kind of response to my early drafts that I wish teachers could take for a model. Sharing my aims for the book, she showed me where and how to shape it toward them. I am grateful to her for faith in a project risky for both of us. Of course, the ultimate responsibility for the accuracy of facts and the quality of judgments is mine.

Introduction

The educational spotlight is on goals, standards, assessments. Since fall 1989, when President George Bush met with the governors of all fifty states to outline a strategy for improving the achievement of American students, energy at the state and national level has been invested in setting national goals and figuring out how to measure progress toward them.

This book is an important component of the national debate, because it describes for people both inside and outside the educational community the kinds of assessment frequently mentioned—“alternative,” “authentic,” “performance-based,” or (my preferred term) “performance assessment.” These terms mean little or nothing to those who quite naturally have taken educational testing for granted as a matter for psychometric specialists. The goal-setting process and the consequent need for valid measurement have thrust educational testing into the foreground, so that everyone concerned with the quality of education not only needs to know what is available, but has a legitimate right to an opinion on it. The President and the governors have made education everyone’s business.

The history of the past two years reveals an interest in education as a national issue that has not been seen in Washington before. Following the “education summit” in Fall 1989, when the President closeted himself with the governors to discuss what they perceive as a national emergency, the National Governors Association announced the national goals for American education in spring 1990. As listed in *America 2000*, they challenged the nation to redirect its efforts so that by the year 2000:

1. All children will start school ready to learn
2. At least 90 percent of high school students will graduate
3. American students will achieve competency in English, mathematics, science, history, and geography at grades 4, 8, and 12; and will be prepared for responsible citizenship, further learning, and productive employment in a modern economy
4. U.S. students will be the first in the world in science and mathematics achievement
5. Every adult American will be literate and able to exercise the rights and responsibilities of citizenship
6. All American schools will be free of drugs and violence

But how do we know whether we're achieving these goals, especially goals 3 and 4, their academic core? Designing indicators of progress became the business of the National Education Goals Panel, which was set up by the National Governors' Association to monitor progress. The Goals Panel appointed advisory committees for each of the six goals, groups which include people whose work is described or words quoted in this book. The Goals Panel was charged with producing the annual Progress Report—the nation's "report card"—a summary of where we are on the way to the six national goals.

When the first Progress Report was published in September 1991, there wasn't much information available on the national level. The report was cobbled together from existing assessments. The framers of the report included the National Assessment of Educational Progress and the number of students taking the College Board's Advanced Placement examinations and their scores—but they did not include the Scholastic Aptitude Test (SAT) nor the American College Test (ACT), nor any norm-referenced achievement tests. The National Assessment of Educational Progress and the Advanced Placement examinations were cited as national information along with statistics on high school course enrollments, public satisfaction with education, and data from international student achievement comparisons, such as those conducted by the International Evaluation of Educational Achievement.

Neither the National Assessment of Educational Progress nor the Advanced Placement examinations are regarded as satisfactory measures of progress toward the goals. The National Assessment takes a

sample of students at certain grades and in certain subjects. When it was established by Congress in 1969, the National Assessment was prohibited from gathering information below the national level—states could not be compared to states, districts to districts, or students to students. In 1990, a pilot comparison of states was permitted, with voluntary participation by states. This resulted in the May 1991 report on the mathematics achievement of eighth graders in thirty-seven states. Obviously it isn't a complete measure of mathematical knowledge and skill.

The Advanced Placement program is an individual student examination, but it is taken only by the most academically advanced students who are intending to apply to prestigious colleges. Clearly this too is an inadequate measure of nationwide educational attainment. But Advanced Placement examinations challenge students to high academic achievement—in fact they are geared to first-year college courses—so they are a measure of how many students are being challenged in which high schools.

Obviously, if the annual Progress Report is to provide meaningful information, it must be based on comprehensive, timely information of the sort that would result from a national examination taken by all students. Such an examination does not exist, but it is being discussed widely, especially by the National Education Goals Panel's advisory groups for goals 3 and 4, the National Council on Education Standards and Testing, and the New Standards Project.

We have already seen where the National Education Goals Panel fits in. The National Council on Education Standards and Testing had its origins in the political background of the education reform movement. The charge was led by the President and the governors—the Senate and the House of Representatives had almost no role. In the middle of 1991, it became obvious that some better measures of educational progress were needed and that they could not be designed without standards. You can't measure how close you've got if there is no clear mark to shoot at. A Senate bill setting up a National Council on Education Standards was discussed in 1990, but got nowhere. Now the idea was revived and both houses of Congress passed legislation to establish the National Council on Standards and Testing. The Council's charge from Congress was to study the "desirability and feasibility" of national standards for American education. It is widely believed that the Council is essentially setting up a national examination system, something that will complement—

perhaps even replace—the National Assessment of Educational Progress, which by its nature has little effect on teaching and learning.

As such bodies do, the Council represents interested parties: there are four members of Congress; U.S. Department of Education administrators; representatives of business, industry, higher education, and the psychometric community; and the whole is chaired by two governors. Their first two public hearings symbolized two opposing ways of setting standards: starting from the top or starting from concrete examples.

Starting from the top involves deciding what kinds of knowledge and skills we want the educational system to produce. At its first public meeting, the National Council therefore asked representatives of the five core academic disciplines—literacy, mathematics, science, history, and geography—to describe what content they expected students to master after twelve years of schooling. This procedure turned out unsatisfactorily. The professional organizations (with the exception of the National Council of Teachers of Mathematics which produced *Curriculum and Evaluation Standards for School Mathematics* in 1989), are fighting among themselves about the definition of their disciplines and what knowledge and skills should be expected from students. The most public dispute is between representatives of social studies teachers' organizations, who believe that history is only one of a number of social studies, and the professional historical associations, who believe students should understand the narrative sweep of history primarily, and social studies will be absorbed along the way. The teachers' professional organizations in science and literacy (reading and writing in English) are equally preoccupied with internal disputes.

A further cause for dissatisfaction with top-down standard-setting is the fragmentation resulting from looking at education within these traditional boundaries. The Council for Basic Education since its founding thirty-five years ago has advocated a vision of an educated person as a productive member of society—one who works, votes, and engages in lifelong learning. The exact amount and nature of mathematics or science or geography courses taken in school should be subordinate to that overall vision.

The second approach to setting standards was exemplified at a subsequent meeting of the National Council, when three people whose work is described in this book explained performance assessment to the members. Tej Pandey described open-ended mathematics

questions used in California statewide assessments (he displayed the “James” question which appears on page 68); Joan Baron demonstrated the real-life problems which make mathematics and science accessible for high school students in Connecticut, and Ross Brewer explained the first statewide assessment by portfolio (collections of student work) in Vermont. Their presentations showed how standards can be described in terms of tasks that students should be able to perform. If students can solve a problem about water use in their own homes, then they have learned important mathematics and can also write to communicate their solutions.

Both approaches to standard-setting are clearly needed. We need a vision to inspire our efforts, but we also need concrete examples of what the vision entails. It is fine to say that all high school graduates should be prepared to vote intelligently, but does that mean knowing the names of candidates running for president, understanding editorials in *The New York Times*, or being able to quote statistics about the exact area of federally protected wetlands in any given state? Such questions translate easily into assessments, so that standards can be approached practically as assessment issues.

That approach is being taken by the New Standards Project, a non-governmental group funded by the MacArthur and Pew Foundations, which is designing a national examination system—an innovation in the United States, although common in many other countries. The New Standards Project has a vision of regions or clusters of states or school districts designing their own examinations, including in them “anchor” tasks. These tasks will be performed by all students taking the examinations, and will be scored nationally. This is the process known as “moderation” in Great Britain, Australia, and some other countries. Scores on the examinations will be equated (“calibrated”) to the anchor tasks, providing comparisons to a national standard. Individual students and their schools will thus know where they stand.

During summer 1991, the New Standards Project established that calibration among different responses to different writing tasks is possible, although a great deal of psychometric sophistication will be needed to make it work for a national examination system. At a working conference attended by more than 350 writing and mathematics teachers and another 80 or so policymakers, the New Standards Project decided on two policies: (1) any state, cluster of states, or regional organization entering into the Project’s national examination system must guarantee the resources to enable every student

to reach the standards exemplified by the tasks; and (2) the tasks themselves must include the “three p’s”—performance, project, and portfolio. You will find examples and explanations of the three p’s in chapters 2 through 6.

This has been a severely reduced summary of the present ferment in American education. I have focussed on goals and standards to demonstrate how central performance assessment is to their attainment. I have not mentioned the President’s America 2000 strategy, or the state versions (“Colorado 2000”) springing up in response; or the New American Schools Development Corporation, or Educate America, or the American Achievement Tests; or the proposals to change the requirements for Chapter 1 compensatory education, so that norm-referenced multiple-choice tests can be replaced with performance assessments for reporting purposes; or the separate standard-setting efforts by the National Academy of Sciences, the U.S. Department of Energy, and the U.S. Department of Labor. It is easy to get confused by the multiplicity of publications, meetings, conferences, and task forces, not to mention their cross-connections.

A consensus is building, however, on the role of assessment as a lever to crank the system up to higher achievement. High standards demand assessments which teachers not only can, but should, teach to, and that students must study for. As you will see from the examples, these assessments are in the early stages of experimentation, but they should not be judged and rejected for lack of qualities which need time to develop. Performance assessment has a vital role in the present educational drama: it will not only chart progress toward the national goals, it will also help us reach them.

Contents

Preface	<i>vii</i>
Introduction	<i>xiii</i>
1 From Testing to Assessment	1
2 Assessing Writing in California, Arizona, and Maryland	27
3 Examples of Performance Assessments in Science and Mathematics	51
4 Getting Students, Parents, and the Community into the Act	79
5 Portfolios	103
6 Teachers and Assessment: Privilege and Responsibility	133
7 How American Education Got into the Testing Trap	165
8 Next Steps	185
Notes	195
Bibliography	207
Index	218