

Educational Assessment of Students

Third Edition



Anthony J. Nitko

6449
NAJ✓

THIRD EDITION

Educational Assessment of Students

ANTHONY J. NITKO

University of Arizona

Merrill
Prentice Hall

Upper Saddle River, New Jersey

Columbus, Ohio

Dedicated to Veronica

Library of Congress Cataloging in Publication Data

Nitko, Anthony J.

Educational assessment of students / Anthony J. Nitko.--3rd ed.
p. cm.

Includes bibliographical references and index.

ISBN 0-13-013708-1

1. Educational tests and measurements. I. Title.

LB3051.N57 2001

371.26--dc21

00-023012

Vice President and Publisher: Jeffery W. Johnston

Executive Editor: Kevin M. Davis

Editorial Assistant: Christina Kalisch

Development Editor: Heather Doyle Fraser

Production Editor: Julie Peters

Production Coordination: Clarinda Publication Services

Design Coordinator: Diane C. Lorenzo

Cover Designer: Allen Bumpus

Cover Photo: UniPhoto

Production Manager: Laura Messerly

Director of Marketing: Kevin Flanagan

Marketing Manager: Amy June

Marketing Services Manager: Krista Groshong

This book was set in Garamond and Arial Narrow by The Clarinda Company. It was printed and bound by The Banta Company. The cover was printed by Phoenix Color Corp.

Earlier edition, entitled *Educational Tests and Measurement: An Introduction*, © 1983 by Harcourt Brace Jovanovich.

Copyright © 2001, 1996 by Prentice-Hall, Inc., Upper Saddle River, New Jersey 07458. All rights reserved. Printed in the United States of America. This publication is protected by Copyright and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. For information regarding permission(s), write to: Rights and Permissions Department.

Merrill
Prentice Hall

9 8 7 6 5 4 3
ISBN 0-13-013708-1

Preface

The goal of *Educational Assessment of Students*, Third Edition, is to help teachers and teachers in training to improve instruction through better assessment of students. It focuses directly on the professional practices of elementary and secondary schoolteachers. It is a core text written for a first course in educational testing and constructing classroom assessments. It serves equally as the textbook for an undergraduate course or a first graduate course in educational assessment. No formal course work in statistics or college mathematics is necessary to understand the text.

In preparing this edition, a new effort was made to make it easy for the reader to apply the material to classroom practice through clear explanations, many practical examples and illustrations, checklists, and step-by-step how-to instructions. As with previous editions, I have written the text from the viewpoint that teaching and assessment should blend together. To help students with real-world experiences, Internet website addresses have been added throughout to help students to quickly link to examples and resources outside the text.

The book covers basic concepts as well as a complete treatment of educational assessment—from developing plans that integrate teaching and assessment; to crafting objective, performance, and portfolio assessments; to evaluating students and discussing evaluations with parents; to interpreting state-mandated tests and standardized achievement tests.

It is important in a first course that students receive a balanced treatment of the topics. Thus, both the strengths and limitations of each assessment technique are explained. Research is cited that supports or refutes assessment and teaching practices.

This edition focuses more clearly on classroom assessment than the previous edition. Each chapter was revised with the goal of making the material more relevant to the practical issues teachers face. For example, mandated state programs have placed more emphasis on teachers aligning their classroom practices and assessments with state content and performance standards. In each chapter, I integrate suggestions for how to accomplish this alignment.

The book treats teachers as professionals. It recognizes that teachers' experiences and judgments are necessary for proper and valid use of educational assessment. I do not hesitate to point out teachers' erroneous judgments and assessment abuses, however, where good lessons can be learned from them. To ensure that the text material is in keeping with the competencies and standards held to be important by the profession, each chapter is keyed to the American Federation of Teachers, the National Council

on Measurement in Education, and the National Education Association *Standards for Teacher Competence in Educational Assessment of Students* (reproduced in Appendix A).

Organization of the Text

As in the previous edition, this book is organized into three parts. Part One provides the foundation for classroom assessment. Part Two centers on how to plan and craft classroom assessments. Part Three discusses how to interpret and use standardized tests.

Each chapter begins with a list of learning targets that the reader should attain. Each learning target is keyed to the *Standards for Teacher Competence in Educational Assessment of Students*. Following the body of each chapter is a summary in list form to help the reader recall information. Each chapter concludes with a list of important concepts and terms, a set of exercises and discussion questions, and a brief annotated list of additional readings. Many of the exercises at the end of the chapters are performance tasks. These require applying and extending the principles taught in the chapter. Readers are often encouraged to compare their exercise results with those of others in the course to obtain greater insight into a topic. Instructors can use these exercises as a basis for in-class activities.

Part I, The Bases for Assessment in the Classroom, contains five chapters presenting the foundations for educational assessment. Chapter 1 discusses classroom decision making and how assessments help to improve it. It provides a brief overview of the assessment landscape, so the reader can see the broader picture of educational assessment before delving into specifics. Chapter 2 discusses goals and learning targets of classroom instruction. The chapter is simplified in this edition by presenting fewer cognitive taxonomies than the previous edition. New to this edition is a discussion of how state standards can be used to develop classroom learning targets for day-to-day teaching. Chapter 3 discusses the validity of assessment results. The chapter has been simplified to focus exclusively on the current conceptions of validity held by the profession. One part of the chapter focuses exclusively on the validity of the results of teacher-crafted assessments, the other on the validity of other assessment results a teacher and a school will use. Chapter 4 describes reliability. The order of presentation of the topics has been improved to make the material flow better for the reader. Chapter 5 discusses teachers', students', and administrators' ethical responsibilities and uses of assessments. It applies the National Council on Measurement in Education's *Code of Professional Responsibilities in Educational*

Measurement (reproduced in Appendix C) as an organizational framework. Also, this chapter describes legal issues regarding assessment and how assessments can be made to validly accommodate mainstreamed learners.

Part II, *Crafting and Using Classroom Assessments*, demonstrates how a teacher uses artistry and functionality to produce useful teaching tools. The ten chapters in this part provide complete coverage of how to craft virtually all types of assessments a teacher may need to use. Chapter 6 has been extensively rewritten to show clearly how teaching, assessing, and grading fit together. There are new examples showing how several different methods of assessment can be integrated into lesson plans, be matched with assessment purpose, and provide formative and summative assessment throughout an instructional unit. Chapter 7 discusses completion and true-false items, including multiple true-false items. Chapter 8 discusses how to craft multiple-choice and matching exercises. In this edition, material on alternative formats of objective items, such as masterlist and tabular matching exercises, have been integrated into this chapter. Chapter 9 focuses on crafting and marking essay items. Chapter 10 is devoted to assessing higher-order thinking skills, including problem solving and critical thinking. The presentation is simplified from the previous edition.

Two chapters are devoted to alternative and performance assessments in schools. Chapter 11 describes the basic concepts and major components of authentic and performance assessments. It surveys the many different types of performance assessments and projects that are currently used in schools. New to this edition is classroom assessment using the Multiple Intelligences Theory framework. Chapter 12 focuses exclusively on crafting and evaluating performance assessments and portfolios. The chapter shows how to craft the tasks, scoring rubrics, and rating scales necessary to produce valid results. The chapter applies the Dimensions of Learning Model as a framework for performance assessment design. Different methods of crafting rubrics are discussed. Chapter 13 focuses entirely on format, informal classroom assessment for purposes of diagnosing why a student is having difficulty learning. It shows how six different diagnostic assessment approaches may be integrated into the teaching process. Chapter 14 has been simplified, but is still about administering assessments, helping students do their best on assessments, and improving assessment tasks through item analysis. Item analysis for open-response items is new in this edition.

Chapter 15 has been extensively rewritten. It is organized in five major sections. Section One provides background for grading including grading and continuous assessment, formative and summative student assessment, and how various stakeholders use grades. Section Two presents various options that schools use to report student progress, including report cards, parent-teacher conferences, and—new to this edition—narrative reports. Section

Three focuses on sensible grading practices that result in valid summative grades for students. The chapter provides teachers with the tools for creating a grading philosophy that is logically consistent with both their teaching approach and their purposes for assessing students. The tools integrate all the planning considerations discussed in earlier chapters. As a result, teachers can use a coherent and sensible grading model. Linking grades to teaching and assessment plans; and discussions of special problems in grading, such as using zeros and the unclear meaning of failure, are new to this edition. Section Five shows how to set grade boundaries and combine scores to produce summative grades.

The text concludes with Part III, *Interpreting and Using Standardized Tests*. Chapter 16 describes different types of standardized achievement tests, how to administer them, and how to use their results. The chapter has been updated to reflect new editions of tests and to include material on how to select a standardized test that complements a state's mandated assessment program. Chapter 17 covers norm-referenced scores and contrasts them with criterion-referenced scores. Various types of scores are explained, emphasizing how a teacher should interpret and explain them to parents. Chapter 18 discusses how to locate and evaluate an appropriate standardized test to use in a school setting. The chapter updates material on how to use the Internet to obtain assessment information. Chapter 19 briefly summarizes the types of scholastic aptitude, vocational interest, and personality assessments a teacher may encounter in student reports. New editions of the tests are included. The chapter includes a discussion of aptitudes and learning, and how the classroom and school environments affect these.

No single introductory course is likely to cover all of these chapters in detail. However, with this text an instructor can choose material to suit the needs of teachers at different levels of professional development and experience. The book provides ample material for both independent study and later reference while teaching.

Supplemental Material

The appendixes are especially important for complete learning of educational assessment concepts and applications. Appendix A reproduces the *Standards for Teacher Competence in Educational Assessment of Students*, to which the learning targets of each chapter are keyed. Appendix B reproduces the *Code of Fair Testing Practices in Education*, which describes the obligations of test developers and test users in straightforward language. Appendix C reproduces the *Code of Professional Responsibilities in Educational Measurement*. Appendix D summarizes and gives examples of using several cognitive, affective, and psychomotor taxonomies of educational learning targets. Appendix E summarizes the general cross-curriculum learning targets derived from the

Dimensions of Learning Model. Appendix F illustrates the assessment of metacognition. Appendix G shows several alternative ways to craft an assessment blueprint for summative evaluation. Appendix H is a brief description of basic statistical concepts such as mean, standard deviation, and correlation. Appendix I shows examples of how to calculate reliability and decision-consisting indexes. Appendix J lists commonly used published tests and cites the volume and page of the *Mental Measurements Yearbooks* in which they are reviewed. Appendix K lists the major test publishers' Internet websites and postal addresses.

An Instructor's Manual provides numerous test items. The test bank is also planned in computerized form.

Acknowledgments

A project of this magnitude requires the help of many persons. I am very much indebted to the reviewers whose critical readings of the chapters in this edition's manuscript contributed greatly to their technical accuracy, readability, and pedagogy: Susan M. Brookhart, Duquesne University; Deborah Brown, West Chester University; Robert Lange, University of Central Florida; Craig Mertler, Bowling Green State University; Susan E. Phillips, Michigan State University; Anthony Truog, University of Wisconsin-Whitewater; Richard Wolf, Teachers College, Columbia University; and Michael J. Young, University of Pittsburgh.

I would also like to thank the reviewers of the second edition: Peter W. Airasian, Boston College; Lawrence M. Aleamoni, University of Arizona; Carol E. Baker, University of Pittsburgh; W. L. Bashaw, University of Georgia; Susan M. Brookhart, Duquesne University; Lee Doebler, University of Montevallo; Betty E. Gridley, Ball State University; Thomas M. Haladyna, Arizona State University; Charles Hughes, Pennsylvania State University; Louise F. Jernigan, Eastern Michigan University; Suzanne Lane, University of Pittsburgh; William P. Moore, University of Kansas; Pamela A. Moss, University of Michigan; Bruce Rogers, University of Northern Iowa; William M. Stallings, Georgia State University; Hoi K. Suen, Pennsylvania State University; James S. Terwilliger, University of Minnesota; Michael S. Trevisan, Washington State University; Kinnard White, University of North Carolina; and David R. Young, SUNY-Cortland.

My students at the School of Education, University of Pittsburgh; the Curriculum Development and Evaluation Centre, Botswana Ministry of Education; Jamaica Ministry of Education; and the Examination Development Center, Indonesia Ministry of Education and Culture used the second edition. They provided insightful feedback and corrections of errors that have greatly improved the usefulness of the text. Francis Amedahe helped classify chapter learning targets and write tests items. Huixing Tang of the Psychological Corporation, Wendy Yen of CTB/McGraw-Hill, Sara Hennings of Houghton Mifflin, James Impara of the Buros Institute of Mental Measure-

ments, and David Frisbie and Robert Forsyth of the Iowa Testing Program were most generous in helping me to locate information about, obtain copies of, and understand many of the standardized test materials described in this book. Kevin Davis, Julie Peters, Holly Jennings, and Christina Kalisch, of Merrill/Prentice Hall, as well as Emily Autumn of Clarinda Publication Services, encouraged and cajoled me at every stage from acquisition through production, and applied their considerable knowledge and skill to producing the book. To all of these persons, and others I have failed to mention, I offer my most sincere thanks and appreciation.

No one deserves more thanks, for this edition as well as past editions, than my best friend and darling wife Veronica. The bulk of the revisions for this edition were made evening and weekends in Dhaka, Bangladesh. Veronica made suggestions for improvement, providing a teacher's perspective. She typed the entire manuscript during the day while I was away on project activities. Her energy, enthusiasm, and constant encouragement sustained me during this revision process. She claims that "this is the last time," but I hope she will continue should there be another edition. To Veronica, then, I lovingly dedicate this book.

DISCOVER THE COMPANION WEBSITE ACCOMPANYING THIS BOOK

The Prentice Hall Companion Website: A Virtual Learning Environment



Technology is a constantly growing and changing aspect of our field that is creating a need for content and resources. To address this emerging need, we have developed an online learning environment for students and professors alike—Companion Websites—to support our textbooks.

In creating a Companion Website, our goal is to build on and enhance what the textbook already offers. For this reason, the content for each user-friendly website is organized by chapter and provides the professor and student with a variety of meaningful resources. Common features of a Companion Website include:

For the Professor—

Every Companion Website integrates **Syllabus Manager™**, an online syllabus creation and management utility.

- **Syllabus Manager™** provides you, the instructor, with an easy, step-by-step process to create and revise syllabi, with direct links into Companion Website and other online content without having to learn HTML.
- Students may logon to your syllabus during any study session. All they need to know is the web address for the Companion Website, and the password you've assigned to your syllabus.

- After you have created a syllabus using **Syllabus Manager™**, students may enter the syllabus for their course section from any point in the Companion Website.
- Class dates are highlighted in white and assignment due dates appear in blue. Clicking on a date, the student is shown the list of activities for the assignment. The activities for each assignment are linked directly to actual content, saving time for students.
- Adding assignments consists of clicking on the desired due date, then filling in the details of the assignment—name of the assignment, instructions, and whether or not it is a one-time or repeating assignment.
- In addition, links to other activities can be created easily. If the activity is online, a URL can be entered in the space provided, and it will be linked automatically in the final syllabus.
- Your completed syllabus is hosted on our servers, allowing convenient updates from any computer on the Internet. Changes you make to your syllabus are immediately available to your students at their next login.

For the Student—

- **Chapter Objectives**—outline key concepts from the text
- **Interactive Self-quizzes**—complete with hints and automatic grading that provide immediate feedback for students

After students submit their answers for the interactive self-quizzes, the Companion Website **Results Reporter** computes a percentage grade, provides a graphic representation of how many questions were answered correctly and incorrectly, and gives a question by question analysis of the quiz. Students are given the option to send their quiz to up to four email addresses (professor, teaching assistant, study partner, etc.).

- **Message Board**—serves as a virtual bulletin board to post—or respond to—questions or comments to a national audience
- **Net Searches**—offer links by key terms from each chapter to related Internet content
- **Web Destinations**—links to www sites that relate to chapter content

To take advantage of these and other resources, please visit the *Educational Assessment of Students* Companion Website at

www.prenhall.com/nitko

ABOUT THE AUTHOR

Anthony J. Nitko is an adjunct professor, Department of Educational Psychology, University of Arizona, and professor emeritus and former chairperson of the Department of Psychology in Education at the University of Pittsburgh. His research interests include curriculum-based criterion-referenced testing, integration of testing and instruction, classroom assessment, and the assessment of knowledge and higher-order thinking skills. His publications include the chapter "Designing Tests That Are Integrated with Instruction" in the third edition of *Educational Measurement*. He co-authored (with C. M. Lindvall) *Measuring Pupil Achievement and Aptitude*, (with T-C Hsu) *Pitt Educational Testing Aids* (PETA) (a package of computer programs for classroom teachers), and (with R. Glaser) the chapter "Measurement in Learning and Instruction" in the second edition of *Educational Measurement*.

Dr. Nitko has been the editor of the journal *Educational Measurement: Issues and Practice*, and was also editor of *d'News*, the AERA Division D newsletter. Some of the journals in which his research has appeared include *American Educational Research Journal*, *Applied Measurement in Education*, *Educational Evaluation and Policy Analysis*, *Educational Measurement: Issues and Practice*, *Educational Technology*, *Journal of Educational Measurement*, and *Research in Developmental Disabilities*.

Dr. Nitko has been a member of several committees of the American Educational Research Association, was elected secretary of AERA Division D, served on committees of the National Council on Measurement in Education, and was elected to the board of directors and as president of the latter. He received Fulbright awards to Malawi and Barbados and has served as a consultant to various government and private agencies in the United States, Bangladesh, Barbados, Botswana, Indonesia, Jamaica, Malawi, Namibia, and Singapore.

Brief Contents

PART I

The Bases for Assessment in the Classroom

- ✓ 1 Classroom Decision Making and Using Assessment 1
- ✓ 2 Describing the Goals and Learning Targets of Instruction 21
- 3 Validity of Assessment Results 36
- 4 Reliability of Assessment Results 62
- 5 Professional Responsibilities, Ethical Behavior, and Legal Requirements in Educational Assessments 79

PART II

Crafting and Using Classroom Assessments

103

- 6 Planning for Integrating Assessment and Instruction 104
- 7 Completion, Constructed-Response, and True-False Items 129
- 8 Multiple-Choice and Matching Exercises 144
- 9 Essay Assessment Tasks 184
- 10 Higher-Order Thinking, Problem Solving, and Critical Thinking 201
- ✓ 11 Performance, Alternative, and Authentic Assessments 240

- 12 Performance Tasks, Portfolios, Rating Scales, and Scoring Rubrics 266

- ✓ 13 Formal and Informal Diagnostic Assessments 293

- 14 Preparing Students to Be Assessed and Using Students' Results to Improve Your Assessments 310

- 15 Evaluating and Grading Student Progress 335

PART III

Interpreting and Using Standardized Tests

371

- 16 Standardized Achievement Tests 372

- 17 Interpreting Norm-Referenced Scores 392

- 18 Finding and Evaluating Published Assessments 422

- 19 Scholastic Aptitude, Career Interests, Attitudes, and Personality Tests 432

APPENDIXES

455

- References 492

- Author Index 503

- Index 506

Contents

PART I THE BASES FOR ASSESSMENT IN THE CLASSROOM 1

✓ Chapter 1 Classroom Decision Making and Using Assessment 2

Teachers' Classroom Decisions 3 ■ Distinctions Among Assessments, Tests, Measurements, and Evaluations 4 ■ Assessment and Educational Decisions About Students 9 ■ Norm-Referenced and Criterion-Referenced Interpretations 13 ■ Additional Ways to Describe Assessments 13 ■ Acquiring the Competence to Assess Students 17

✓ Chapter 2 Describing the Goals and Learning Targets of Instruction 21

Importance of Specifying Objectives 21 ■ Taxonomies of Learning Targets 26 ■ Cognitive Domain Taxonomies 26 ■ Sources for Locating Learning Targets 30 ■ Evaluating the Learning Targets of a Course or Unit 30 ■ Desirable Qualities of Specific Learning Targets 30 ■ Making Sure Assessment Tasks Match Learning Targets 32

Chapter 3 Validity of Assessment Results 36

General Nature of Validity 36 ■ Validity of Teacher-Made Classroom Assessment Results 38 ■ Validity of Extra-Classroom Assessments 43 ■ Categories of Validity Evidence 44 ■ Combining Evidence for Validity Judgments 58

Chapter 4 Reliability of Assessment Results 62

General Nature of Reliability 62 ■ Causes of Measurement Error or Inconsistency 63 ■ Types of Reliability Coefficients 65 ■ Obtained Scores, True Scores, and Error Scores 70 ■ Standard Error of Measurement 71 ■ Additional Factors Affecting Reliability and SEM 73 ■ Reliability of Mastery Decisions 74 ■ Desired Levels of Reliability 76

Chapter 5 Professional Responsibilities, Ethical Behavior, and Legal Requirements in Educational Assessments 79

A Teacher's Professional Responsibilities in Assessment 80 ■ Students' Rights as Test Takers 87 ■ Students' Responsibilities 88 ■ Secrecy, Access, Privacy, Confidentiality, and the Teacher 89 ■ Accommodations for Students with Disabilities 90 ■ Testing Challenged in Court 92 ■ Bias in Educational Assessment 94

PART II CRAFTING AND USING CLASSROOM ASSESSMENTS 103

✓ Chapter 6 Planning for Integrating Assessment and Instruction 104

The Seamless Fabric of Assessment and Instruction 104 ■ Are You Assessing for Formative or Summative Purposes? 105 ■ Assessment Planning for a Marking Period and for One Instructional Unit 107 ■ Pretesting in Order to Plan Your Teaching 111 ■ Crafting a Plan for One Summative Assessment 112 ■ Blueprints for Student-Centered Assessment 114 ■ Criteria for Improving the Validity of Summative Assessments 114 ■ What Range of Assessment Options Is Available? 118 ■ Validity of Different Classroom Assessment Options 124

Chapter 7 Completion, Constructed-Response, and True-False Items 129

Knowledge and Skills to Implement Assessment Plans 130 ■ Three Fundamental Principles for Crafting Assessments 130 ■ Short-Answer and Completion Items 131 ■ True-False Items 135

Chapter 8 Multiple-Choice and Matching Exercises 144

Multiple-Choice Item Format 145 ■ Considerations Before Crafting Items 146 ■ Advantages and Criticisms of Multiple-Choice Items 148 ■ When Not to Use Multiple-Choice Items 149 ■ Crafting Basic Multiple-Choice Items 149 ■ A Checklist for Evaluating Multiple-Choice Items 160 ■ Crafting Alternative Varieties of Multiple-Choice Items 161 ■ Greater-Less-Same Items 161 ■ Best-Answer Items 163 ■ Experiment-Interpretation Items 164 ■ Statement-and-Comment Items 167 ■ Matching Exercise Format 169 ■ Advantages and Criticisms of Matching Exercises 170 ■ Crafting Matching Exercises 171 ■ Crafting Alternative Varieties of Matching Exercises 173 ■ Masterlist (Keylist) Items 173 ■ Tabular (Matrix) Items 176

Chapter 9 Essay Assessment Tasks 184

Varieties of Essay Items 185 ■ Usefulness of Essay Assessments 187 ■ Constructing Essay Items 189 ■ Optional Questions 193 ■ Scoring Essay Assessments 194

Chapter 10 Higher-Order Thinking, Problem Solving, and Critical Thinking 201

Assessing Higher-Order Thinking 202 ■ Context-Dependent Item Sets 202 ■ Concepts and Concept Learning 203 ■ Assessing Concrete Concept Learning 204 ■ Assessing Concept Understanding at a Deeper Level 205 ■ Assessing Defined Concept Learning 206 ■ Using Rules in Thinking 207 ■ Assessing Comprehension of Rules and Principles 208 ■ Problem Solving 210 ■ Assessing Problem-Solving Skills 212 ■ Other Promising Approaches for Assessing Problem Solving 215 ■ Critical Thinking 215 ■ Assessing Dispositions Toward Critical Thinking 216 ■ Assessing Critical Thinking Abilities 216 ■ Other Specific Higher-Order Skills and Abilities 227 ■ Enhanced Multiple-Choice Items 235

✓ Chapter 11 Performance, Alternative, and Authentic Assessments 240

Performance Assessment and Teaching 240 ■ Authentic Assessments 244 ■ Multiple Intelligences Perspective on Performance Assessments 245 ■ Types of Performance Assessments 247 ■ Advantages and Criticisms 261

Chapter 12 Performance Tasks, Portfolios, Rating Scales, and Scoring Rubrics 266

Stages in Crafting Performance Tasks 266 ■ Stage One: Being Clear about the Performance to Assess 267 ■ Stage Two: Crafting Performance Tasks 269 ■ Stage Three: Rubrics and Assessment Results 273 ■ Crafting Portfolios 287

Chapter 13 Formal and Informal Diagnostic Assessments 293

Diagnostic Assessment 293 ■ Integrating Teaching, Diagnosis, and Assessment 295 ■ Approach 1: Profiling Content Strengths and Weaknesses 296 ■ Approach 2: Identifying Prerequisite Deficits 297 ■ Approach 3: Identifying Objectives Not Mastered 299 ■ Approach 4: Identifying Students' Errors 301 ■ Approach 5: Identifying Student Knowledge Structures 302 ■ Approach 6: Identifying Competencies for Solving Word Problems 305 ■ Interviewing Students 307

Chapter 14 Preparing Students to Be Assessed and Using Students' Results to Improve Your Assessments 310

- Section I. Getting Students Ready 311 ■ Preparing Students for the Assessment 311 ■ Minimum Assessment-Taking Skills 312 ■ Test-Wiseness 312 ■ Advice About Changing Answers 313
- Test Anxiety 313 ■ Section II. Putting the Assessment Together 316 ■ Assessment Format and Appearance 316 ■ Preparations for Scoring the Assessment 317 ■ Correction for Guessing 318 ■ Section III. Using Students' Responses to Improve Assessments 319 ■ Item Analysis for Classroom Assessments 319 ■ Item Difficulty Index 323 ■ Item Discrimination Index 325
- Improving Multiple-Choice Item Quality 326 ■ Selecting Test Items 327 ■ Using Computers for Test Assembly and Item Analyses 329

Chapter 15 Evaluating and Grading Student Progress 335

- Section I. Background 336 ■ What Are Your Attitudes Toward Marks and Grades? 336 ■ Continuous Assessment and Grading 336 ■ How People Perceive and Use Grades 337 ■ Criticisms of Grades and Marks 340 ■ Section II. Official Reports of Student Progress 341 ■ Student Progress Reporting Methods 341 ■ Crafting a Multiple Reporting System 348 ■ Section III. Choosing a Grading Model 348 ■ Basic Concepts of Grading Frameworks for Classroom Evaluation 349 ■ Choosing Your Grading Model 352 ■ Section IV. Sensible Grading Practices 354 ■ Types of Behavior to Assess 354 ■ Consistent Grading throughout the Marking Period 355 ■ Components of a Grade 356 ■ Grade Boundaries 356 ■ Failure Grades 357 ■ Section V. Techniques for Setting Grade Boundaries and Combining Scores 359 ■ Assigning Norm-Referenced Letter Grades 359 ■ Assigning Criterion-Referenced Letter Grades 362 ■ Gradebook Computer Programs 365

PART III INTERPRETING AND USING STANDARDIZED TESTS 371

Chapter 16 Standardized Achievement Tests 372

- Types of Tests 373 ■ Standardized Survey Batteries 375 ■ Nonstandardized Achievement Tests 380 ■ State-Mandated Tests 380 ■ Using Standardized Test Results 381 ■ Multilevel Survey Batteries 384 ■ Choosing Standardized Tests 385 ■ Administering Standardized Tests 387

Chapter 17 Interpreting Norm-Referenced Scores 392

- Two Referencing Frameworks 393 ■ Using Norms 394 ■ Overview of Norm-Referenced Scores 397 ■ Normal Distributions 402 ■ Normalized Standard Scores 405 ■ Extended Normalized Standard Score Scales 408 ■ Grade-Equivalent Scores 409 ■ Comparison of Various Norm-Referenced Scores 416 ■ General Guidelines for Score Interpretation 417 ■ Interpreting Scores to Parents 418

Chapter 18 Finding and Evaluating Published Assessments 422

- Information About Standardized Tests 423 ■ Evaluating and Selecting a Test 426 ■ How a Standardized Test Is Developed 429 ■ Qualifications for Test Purchase 429

Chapter 19 Scholastic Aptitude, Career Interests, Attitudes, and Personality Tests**432**

- Aptitudes for Learning 433 ■ Group Tests of Scholastic Aptitude 435 ■ Group Tests of Specific
 Aptitudes 441 ■ Individually Administered Tests of General Scholastic Aptitude 442 ■ Assessing
 Adaptive Behavior 446 ■ Assessing Vocational Career Interests 447 ■ Measuring Attitude 450
 ■ Assessing Personality Dimensions 450

APPENDIXES

A Standards for Teacher Competence in Educational Assessment of Students	455
B Code of Fair Testing Practices in Education	457
C Code of Professional Responsibilities in Educational Measurement	459
D Summaries of Taxonomies of Educational Objectives: Cognitive, Affective, and Psychomotor Domains	463
E Categories of Learning Targets Derived from the Dimensions of Learning Model	468
F Assessment of Metacognition	470
G Examples of Alternative Blueprints for a Summative Unit Assessment	472
H Basic Statistical Concepts	476
I Computational Procedures for Various Reliability Coefficients	485
J A Limited List of Published Tests	489
K List of Test Publishers and Their WebSites	491
References	492
Author Index	503
Subject Index	506

PART I

The Bases for Assessment in the Classroom

- 1 Classroom Decision Making and Using Assessment
- 2 Describing the Goals and Learning Targets of Instruction
- 3 Validity of Assessment Results
- 4 Reliability of Assessment Results
- 5 Professional Responsibilities, Ethical Behavior, and Legal Requirements in Educational Assessments

CHAPTER 1

Classroom Decision Making and Using Assessment

Learning Targets¹

After studying this chapter, you should be able to:

1. Give examples of teaching management decisions you need to make before, during, and after instruction. [4, 1, 2]
2. Give examples of assessment procedures that provide you with useful information for making the decisions you stated in Target 1. [1, 2, 3, 4]
3. Define and explain the relationships among teaching, assessment, testing, measurement, and evaluation. [6, 1]
4. Explain the principles that guide you in the process of selecting, developing, and using educationally meaningful assessments. [1, 3]
5. Describe and explain the wide range of educational decisions for which educators need quality assessments. [1, 6]
6. Distinguish between formative and summative evaluation. [1, 4]
7. Describe the type(s) of information assessments provide for making decisions about diagnostic/remedial actions, feedback to students/parents, feedback to teachers, motivating students, and assigning grades to students. [1, 3, 4]
8. Distinguish among selection decisions, placement decisions, classification decisions, counseling and guidance decisions, and credentialing/certification decisions. [1, 3, 6]
9. Explain criterion-referencing and norm-referencing and how they complement one another in educational decision making. [1, 4, 6]
10. Use technical terminology correctly to describe a given assessment procedure in terms of the kinds of items it uses, how student performance is scored, degree of standardization, administrative conditions, language emphasis of the scoring, emphasis on speed of responding, the basis for interpreting scores, and what attribute is measured. [6, 4, 3]
11. Explain how each of the Standards for Teacher Competence in Educational Assessment of Students is related to your teaching activities before, during, and after instruction. [6, 1, 7]
12. Explain how each of the terms and concepts listed at the end of this chapter apply to educational assessments. [6]

It is almost impossible for you to have attended school without having been exposed to a wide variety of educational and psychological assessment procedures. The fact that you are reading this book for a testing and measurement course places you not only among test takers, but among the successful test takers. Think for a few minutes: How many ways have

¹The Learning Targets that begin each chapter are cross-referenced to the American Federation of Teachers, National Council on Measurement in Education, and National Education Association (1990), *Standards for Teacher Competence in Educational Assessment of Students* (see Appendix A). The numbers that follow each learning target statement refer to one or more of the standards. The order of the standards represents the closeness of the match to the standards: The closest match is listed first. Francis Amedahe helped to classify the learning targets.

you been assessed in your life? When did your assessment experiences begin? Consider this example:

Meghan's educational assessment began in kindergarten with an interview and an observation. The state in which she lived had no mandatory kindergarten requirement. On registration day, Meghan and her mother came to school and were briefly interviewed. Meghan's cognitive and social-emotional skills were rated by a teacher. Her development was judged normal, and she attended kindergarten.

During the year she experienced difficulty in paying attention to the teacher and participating in group activities, although she was neither aggressive nor hostile. She was given a "readiness test" at the end of kindergarten and performed as an average child. Her teacher recommended that she continue on to first grade, but her parents balked: They didn't think that she was ready.

They took her to a child guidance clinic and requested further psychological assessment. The clinical psychologist administered an individual intelligence test and a "projective test" in which she was asked to tell a story about what was happening in each of a set of pictures. The psychologist interviewed her, her parents, and her teacher. She was described by the psychologist as normal, both in cognitive ability and in social-emotional development.

Her parents withdrew her from the school she was attending and placed her in another school to repeat kindergarten. Later, they reported that whereas her first experience was difficult for her, her second kindergarten year was a great success. In their view, a teacher who was particularly sensitive to Meghan's needs helped her cognitive development to proceed rapidly. By the end of the year she had also become more confident in herself and regularly participated in group activities.

In the situation just illustrated, assessments provided information. Different persons evaluated the information and reached different conclusions. Decisions needed to be and were made, however. Perhaps different decisions could have been made based on the same information.

This brief anecdote shows assessments being used rather early in the person's life. Most of us recall more easily the assessments applied to us later in our lives, as older children, and as adults. You may not even associate the term *assessment* with Meghan's interviews. Yet, as shall be explained later, the interview is included in the broad definition of assessments, because the basic principles of assessment apply as well to it as to other, more familiar procedures.

Meghan's situation also illustrates that assessment results can contribute to a decision, but the results may not be interpreted in the same way by everyone concerned. Although it may appear that Meghan's parents were right in having her repeat kindergarten, there is no way of ascertaining what would have happened if she had gone straight to first grade, because she didn't.

Decisions involve the use of various kinds of information. Sometimes test scores play a major role; at other

times, less formal assessments play a more dominant role. This book examines a variety of decisions for which assessments are used in education, especially in the classroom. Each time, it identifies the basic principles that relate to the evaluation and use of assessment information. It emphasizes basic principles rather than prescriptions to follow blindly, even though you will find some of those too. However, if you understand the basic principles, you should be able to determine when those and other prescriptions are inappropriate.

TEACHERS' CLASSROOM DECISIONS

Teaching and learning require you to constantly gather information and make decisions. Teachers often need to make decisions about students at the rate of one every 2 to 3 minutes (Shavelson & Stern, 1981). Sound teaching decisions require sound information. Sound assessment procedures gather sound information. Researchers estimate that teachers may spend from one third to one half of their time in assessment-related activities (Stiggins, Conklin, & Associates, 1992). A few examples of questions you must answer when making teaching decisions follow. Examples of assessment procedures that may give you useful information for making the decisions are listed in parentheses.

Decisions Before Beginning Teaching

1. What content do I need to cover during the next day, week, month, marking period, etc.? (*Review the curriculum, the syllabus, the textbook, and the formal tests your students will need to pass.*)
2. What abilities (cultural background factors, interests, skills, etc.) of my students do I need to take into account as I plan my teaching activities? (*Informal observation of the students during class; conversations with students' previous teachers; scholastic aptitude test results' previous teachers; scholastic aptitude test results; students' past grades and standardized test results; knowledge of the student's personal family circumstances.*)
3. What materials are appropriate for me to use with this group of students? (*Informal observation of students' motivations, interests, beliefs, and experience with the content you will teach; informal observations about students' attitudes toward learning the topics at hand; results from pretests, previous teacher's evaluations, and standardized achievement tests.*)
4. With what learning activities will my students and I need to be engaged as I teach the lesson (unit, course)? (*Review the types of activities I used previously that stimulated the interests of my students; analyze the sequence of the learning activities students will follow; review the student achievement that resulted when the activities were used with other students.*)

5. What learning targets do I want my students to achieve as a result of my teaching? (*Review statements of goals and learning objectives; review test questions students should be able to answer; review performances and thinking skills students should be able to demonstrate after learning.*)

6. How should I organize and arrange the students in the class for the upcoming lessons and activities? (*Informal observation of students with special learning and social needs; informal observation and recall of students' behavior during previous learning activities; information about what classroom arrangements worked best in the past when my students were learning similar targets.*)

Decisions During Teaching

1. Is my lesson going well? Are students catching on (i.e., learning)? (*Observations of students during learning activities; student responses to questions I have asked them; observing student interactions.*)

2. What should I do to make this lesson (activity) work better? (*Diagnosis of the types of errors students made or erroneous thinking students are using; searching my memory for alternative ways to teach the material; identifying which students are not participating or are acting inappropriately.*)

3. What feedback should I give each student about how well he or she is learning? (*Informal observation and experience on the amount and type of praise different students require; information about the quality of each student's learning of the intended target performance; homework and quiz results; interviewing students.*)

4. Are my students ready to move to the next activity in the learning sequence? (*Informal observation and checking students' completed work; questioning students about their understanding; students' homework, quiz, and test results.*)

Decisions After a Teaching Segment

1. How well are my students achieving the short- and long-term instructional targets? (*Classroom tests, projects, observations, interviews with students, standardized test results.*)

2. What strengths and weaknesses will I report to each student and to his or her guardian or parent? (*Observations of each student's classroom participation; review each student's homework results; review each student's standardized achievement and scholastic aptitude test results; review information about a student's personal family circumstances.*)

3. What grade should I give each student for the lesson or unit, marking period, or course? (*Results from classroom learning activities, quizzes, tests, class projects, papers, labs, etc.; informal observation about how well the student has attained the intended learning targets; knowledge of the possible*

short- and long-term consequences to the student of reporting a particular grade.)

4. How effectively did I teach this material to the students? (*Summaries of the class's performance on the important instructional targets; summaries of the class's performance on selected questions on standardized tests; summaries of how well the students liked the activities and lesson materials.*)

5. How effective are the curriculum and materials I used? (*Summaries of informal observations of students' interests and reactions to the learning activities and materials; summaries of the class's achievement on classroom tests that match the curriculum; summaries of several classes' performance on selected areas of standardized tests.*)

These lists of questions are not exhaustive; you may wish to list several others. Later in this chapter we discuss several other educational decisions that need to be made by and for students. However, the preceding examples do illustrate that your teaching decisions require you to use many different types of information. Further, they illustrate that the exact type of information you need varies greatly from one teaching situation to the next. Remember that you cannot expect to make good decisions as you teach unless you have good quality information on which to base these decisions.

DISTINCTIONS AMONG ASSESSMENTS, TESTS, MEASUREMENTS, AND EVALUATIONS

The general public often uses the terms *assessment*, *test*, *measurement*, and *evaluation* interchangeably, but it is important for you to distinguish among them. The meanings of the terms, as applied to situations in schools, are explained in the following paragraphs.

Assessments

Assessment is a broad term defined as a process for obtaining information that is used for making decisions about students, curricula and programs, and educational policy (see American Federation of Teachers, National Council on Measurement in Education, and National Education Association, 1990). Decisions about students include managing classroom instruction, placing students into different types of educational programs, assigning them to appropriate categories, guiding and counseling them, selecting them for educational opportunities, and credentialing and certifying their competence. Decisions about curricula and programs include decisions about their effectiveness (summative evaluations) and about ways to improve them (formative evaluations). Decisions about educational policy are made at the local school

district level, the state level, and the national level. Figure 1.1 illustrates some of the decisions and subdecisions for which educational assessments provide information.

When we say we are “assessing a student’s competence,” we mean we are collecting information to help us decide the degree to which the student has achieved the learning targets. A large number of assessment techniques may be used to collect this information. These include formal and informal observations of a student; paper-and-pencil tests; a student’s performance on homework, lab work, research papers, projects, and during oral questioning; and analyses of a students’ records.

Guidelines for Selecting and Using Classroom Assessments

Although *assessment* is a broad and comprehensive term, you should not think of assessment as simply referring to a collection of techniques. Neither should you think that every information-gathering activity is assessment. Remember that assessment is a process for obtaining information for making a particular educational decision. Because you should focus your assessment activities on the information you need to make particular educational decisions, you need to become competent in selecting and using assessments. Here is a set of guiding principles that you should follow to select and use educational assessments meaningfully.

1. *Be clear about the learning target you want to assess.* Before you can assess a student, you must know the kind(s) of student knowledge, skill(s), and performance(s) about which you need information. The knowledge, skills, and performances you want students to learn are sometimes called *achievement targets*. The more clearly you are able to specify the learning targets, the better you will be able to select the appropriate assessment techniques.

2. *Be sure that the assessment technique(s) you select actually match the learning target.* “Do we want to evaluate students’ problem-posing and problem-solving in mathematics? Experimental research in science? Speaking, listening, and facilitating a discussion? Doing document-based historical inquiry? Thoroughly revising a piece of imaginative writing until it ‘works’ for the reader? Then let our assessment(s) be built out of such exemplary intellectual challenges” (Wiggins, 1990, p. 1). The assessment techniques selected should be as practical and efficient to use as possible, but practicality and efficiency should not be the overriding considerations.

3. *Be sure that the selected assessment techniques serve the needs of the learners.* Proper assessment and evaluation show students concrete examples of what they are expected to do with their learning. Assessment techniques should provide learners with opportunities for determin-

ing specifically what they have achieved and specifically what needs to be done to improve their performance. Therefore, you should select assessment techniques that provide meaningful feedback to the learners about how closely they have approximated the learning targets. Good assessment is good instruction.

4. *Whenever possible, be sure to use multiple indicators of performance for each learning target.* One format of assessment (such as short-answer questions or matching exercises) provides an incomplete picture of what a student has learned. Because one assessment format tends to emphasize only one aspect of a complex learning target, it typically underrepresents that learning target. Using triangulation or multiple modalities of assessment usually enhances the validity of your assessments. Matching exercises, for example, emphasize recall and recognition of factual information; essay questions emphasize organizing ideas and writing skill under the pressure of time limits; and a month-long project emphasizes freely using resources, research, and more thorough analyses of the topic. All three of these assessment techniques may be needed to ascertain the extent to which a student has achieved a given learning target.

5. *Be sure that when you interpret the results of assessments you take their limitations into account.* Although Guiding Principle 2 calls for increasing the authenticity or meaningfulness of the assessment techniques, assessments that occur in schools cannot completely reproduce those things we want students to learn in “real life.” The information we obtain, even from multiple assessments, is only a *sample* of a student’s attainment of a learning target. Because of this, information from assessment contains *sampling error*. Also, factors such as a student’s physical and emotional conditions further limit the extent to which we can obtain truly accurate information. Teachers, and others, must make decisions nevertheless. The decisions, however, must keep these limitations in mind.

Tests

Testing Individuals A **test** is a concept narrower than assessment. It is defined as an instrument or systematic procedure for observing and describing one or more characteristics of a student using either a numerical scale or a classification scheme. In schools, we usually think of a test as a paper-and-pencil instrument with a series of questions that students must answer. These tests are usually scored by adding together the “points” a student earned on each question. Thus, they describe the student using a numerical scale. Similarly, a preschool child’s cognitive development could be observed by using the *Wechsler Pre-School and Primary Scale of Intelligence* (see Chapter 19) and described as having a percentile rank of 50 (see Chapter 17). Other systematic observation procedures such as the play