

EDUCATIONAL EVALUATION AND MEASUREMENT

COMPETENCIES FOR ANALYSIS
AND APPLICATION

SECOND EDITION



L.R. GAY

**EDUCATIONAL
EVALUATION AND
MEASUREMENT
COMPETENCIES FOR ANALYSIS
AND APPLICATION
SECOND EDITION**

L.R. Gay

Florida International University

Charles E. Merrill Publishing Company
A Bell & Howell Company
Columbus Toronto London Sydney

Published by Charles E. Merrill Publishing Co.
A Bell & Howell Company
Columbus, Ohio 43216

This book was set in Palatino and Helvetica.
Production Coordinator: Megan Rowe
Cover Designer: Cathy Watterson

Photo Credits:

- p. 2—Courtesy of Twentieth Century-Fox. Copyright © 1979 Twentieth Century-Fox Film Corp. All rights reserved.
p. 54—*New York Daily News*.
pp. 101, 103—Norman Rockwell, *Saturday Evening Post*.
p. 152—*St. Jerome Studying in His Cell* by George de la Tour. The Louvre.
p. 205—*Faith and Confidence* by William C. Beall.
p. 298—From the television series "THE INCREDIBLE HULK." Copyrighted by Universal City Studios, Inc. THE INCREDIBLE HULK is a trademark of the Marvel Comics Group. Used with permission.
p. 366—*The Gleaners* by Jean Francois Millet. The Louvre.

Copyright © 1985, 1980, by Bell & Howell Company. All rights reserved. No part of this book may be reproduced in any form, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Library of Congress Catalog Card Number: 84-62767
International Standard Book Number: 0-675-20442-9
Printed in the United States of America

3 4 5 6 7 8 9 10 11 12 13 14 K 4 3 2 1 0

To my sisters

Pat, Dot, and Debbie Hines

Preface

Traditionally, schools and colleges of education have acknowledged that professionals in all areas of education need to possess measurement knowledge and skills. A measurement course may be a basic requirement for undergraduate education majors, or it may be a requirement for a graduate degree, but at some point most education students must take such a course. Increasingly, however, the need for instruction in the broader field of *evaluation* is being recognized. Some schools and colleges are attempting to meet this need by offering a separate course in evaluation. Others are supplementing existing measurement or research courses with evaluation topics.

The philosophy which guided the development of this text is that measurement and evaluation are not separate disciplines. Measurement is an integral part of the evaluation process. The purpose of this text is to present measurement concepts within the framework of contemporary evaluation. This book is designed primarily for use in either an introductory-level graduate course or a senior-level undergraduate course. Since the topic coverage is relatively comprehensive, however, this text may easily be adapted for use in either a more advanced graduate course or a lower-level undergraduate course.

The instructional philosophy behind this text is that an introductory course in measurement and evaluation should be skill oriented rather than knowledge oriented, and application oriented rather than theory oriented. Thus, the purpose of this book is *not* simply to have students become familiar with evaluation procedures or acquire a body of knowledge. The purpose of this book is *not* to mystify students with theoretical and statistical jargon. The purpose of this book is *not* simply to have students acquire in-depth understanding of theory. The purpose of this book is to help students acquire the skills and knowledge needed to be a competent consumer and producer of measurement and evaluation tools and techniques. The emphasis is not on what the student knows but rather on what the student can do with what he or she knows. Expertise involves more than the acquisition of skills and knowledge; through experience one acquires insights, intuitions, and strategies related to measurement and evaluation. Experience has meaning, however, only if it is related to a foundation of basic skills and knowledge. Another basic assumption of this text is that there is considerable overlap in the skills and knowledge needed to be a competent consumer of measurement and evaluation tools and techniques and those needed to be a competent producer of same. A person is in a much better position to assess the efforts of others *after* she or he has personally performed the major tasks involved. The instructional strategy of the text, then, is to help students ac-

quire a knowledge of concepts and procedures *and* to involve them in processes of measurement and evaluation.

Two of the most important points about evaluation are made in part 1. First, although the emphasis of the text is on student evaluation, the basic evaluation process is the same regardless of what is being evaluated, and the basic concepts of measurement are the same regardless of what is being measured. Many people believe that program or project evaluation and evaluation of instruction, for example, involve entirely different processes. The current proliferation of "different" evaluation models has reinforced this assumption. All systems and models for evaluation involve the same essential components, namely: specification of goals and objectives; selection and/or development of measurement tools; delineation and execution of strategies for obtainment; and analysis and interpretation of results. Second, evaluation should not be an afterthought; it should be planned for at the beginning of an endeavor, not at the end.

In part 1, each student identifies and describes an evaluation situation of interest which has relevance to his or her professional life. Since the majority of students in an introductory measurement course are teachers, evaluation of student performance is most often selected. In each succeeding part, the measurement student simulates the procedures involved in the planning, process, and product phases of the evaluation. Specifically, the student specifies objectives (part 2), specifies the measurement technique most appropriate for determining achievement of the objectives (part 3), selects standardized instruments (part 4), designs and develops an instrument and performs an item analysis (part 5), analyzes and interprets the results of testing (part 6), and reports results (part 7).

This book is more than just a textbook to be read as part of a course; it is actually a total instructional system which includes stated objectives, or competencies, instruction, and procedures for assessment of each competency. The system emphasizes demonstration of skills and individualization within a basic structure. The format for each part is essentially the same. Following an introduction, the task to be performed for that part is described. Tasks require students to show that they can perform particular evaluation functions. Since each student works with a different set of objectives (as specified in part 2), each student demonstrates the competency required by a task as it applies to her or his objectives. Each chapter within a part begins with a list of student performance outcomes. These outcomes entail knowledge and skills which should facilitate students' ability to perform the task for a given part. In many instances, these expected outcomes may be assessed by using written exercises submitted by students or by using criterion-referenced tests, whichever the instructor prefers. For some outcomes the first option is clearly preferable. Each part is directed toward the accomplishment of a task.

Discussion of concepts and procedures in the text is as simple and straightforward as possible. Whenever feasible, procedures are presented as a series of steps, and concepts are illustrated with examples. In a number of cases, relatively complex topics beyond the scope of the text are presented at a very elementary level, and students are directed to other sources for in-depth discussion. There is also a degree of intentional repetition. A number of concepts are discussed at different points in the text in different contexts and from different perspectives. Also, at the risk of eliciting more than a few groans, I have sprinkled the text with humorous comments and notes. Each part includes a rather lengthy summary with headings and subheadings which directly parallel those in the chapters. The summaries are designed to facilitate both review and location of topics in the preceding chapters. Following the summary, exercises are presented which allow students to test their comprehension and ability to apply concepts and skills presented in the part. Each part concludes with the performance criteria associated with the part task, and a model example of task performance based on work submitted by a former measurement and evaluation student.

The instructor's manual which accompanies this text contains four sections. The first section describes various ways the text may be used and outcomes assessed. The second section includes, for each part in the text, suggested instructional strategies, discussion questions, and supplementary material on text topics or related topics. The third section provides additional examples of task performance. The fourth section includes nearly 200 fieldtested test items, organized by item type and part, which are classified as mastery items or discriminating items. In combination, the text and instructor's manual facilitate implementation of a variety of instructional approaches.

The second edition reflects input from a number of professors who are current or potential users of the text. Their primary suggestion for text revision was to reduce its length by deleting or condensing 1) material typically found in a research text (e.g., questionnaire construction), and 2) material related to in-depth study of evaluation topics (e.g., needs assessment). Their rationale was that most of their students are teachers, and that the text should consequently concentrate on classroom measurement and evaluation. There were, however, a significant number of professors who felt strongly that coverage of the above-described topics should not be excluded or reduced. Their rationale was that their students, teachers or not, are involved, directly or indirectly, in other types of evaluation. Subsequently, taking all views into consideration, coverage of certain topics was deleted, condensed, and/or placed in the instructor's manual. Except for a brief overview, for example, material on survey instruments and evaluation designs has been deleted. Coverage of statistical concepts has been significantly reduced by deletion of material related to inferential statistics. Discussion of topics such as curriculum evaluation and needs assessment has been limited in the text to presentation of basic concepts, and deleted material has been placed in the instructor's manual. Topics such as goal-free evaluation and instructional strategies and variables have been totally removed from the text and placed in the instructor's manual.

Other content changes were made to either expand or update topic coverage. The discussion of standardized criterion-referenced tests, for example, has been lengthened and includes a detailed description of one such system. Coverage of topics such as the National Assessment of Educational Progress (NAEP) program and minimum competency testing has been revised to reflect their current status.

The major organizational change has been the elimination of the student manual which accompanied the first edition. The rationale for this decision was twofold. First, it was felt that it was more convenient, and perhaps sounder from a learning perspective, to have exercises and task examples included in the text immediately following related material. Second, a number of professors do not wish to place an additional financial burden on students by requiring them to purchase a text *and* a manual. Therefore, all exercises were placed in the text itself, as well as one example of expected performance for each task. As mentioned previously, additional task examples were moved to the instructor's manual.

I express my sincere appreciation to all the EDF 5430 students whose feedback and performance stimulated constant revision and refinement of the content and organization of this text. I also wish to thank Professors Charles Dzuiban, Larry Henriksen, David Krus, William P. McDougall, David Payne, Darrell Sabers, Terry Schurr, Patricia Simun, Roy Sommerfield, and Wilber Williams for their conscientious and helpful reviews of portions of drafts of the first and second editions of the text. I especially thank Professor Joan Michael who carefully reviewed the entire manuscript. Thanks are also extended to Professors Mervin Lynch, Carlton Lemkuhl, and Al Kuolmos for their participation in class testing of text content and organization.

For students only

If you are anything like students I have taught in the past, you are quite a diverse group. Many of you are teachers, some of you are school administrators or counselors, and some of you are involved with state or local education agencies. A few of you are not associated with the public schools at all, but are involved in fields such as nursing, banking, and homemaking. Most of you are using this text in a course required (Ugh!) for a degree.

The fact that you are in the course indicates that *somebody* thought it involved concepts important for you to know. While there is general agreement that measurement and evaluation skills are important, there is considerable disagreement about exactly what you need to know or should be able to do. Your own feelings about what you need and want to know also vary widely. Regardless of this, all of you are involved in evaluation activities every day, whether you realize it or not. Each concept presented in this book represents an aspect of evaluation in which you are involved personally and professionally, directly or indirectly, as an interested participant in or as an interested consumer of evaluation. If you are a teacher, for example, you may see the value of knowing how to develop a classroom test, but may not see the need for being familiar with standardized tests; after all, *you* never give any. As a teacher, however, you are really at the center of all educational evaluation activities. No matter at what level an activity occurs, ultimately you and your students are the participants. Even if you do not administer standardized tests, your students take them and you receive the results.

Years of classroom experience have shown me that students in a measurement and evaluation course *want* to know a lot of things that other people don't think they *need* to know. Teachers, for example, want to be informed about concepts such as competency testing and standardized testing. There would probably be a lot less resistance to many of the evaluation activities that go on in the schools, for example, if all participants were "let in on it," that is, if they had a general understanding of the purpose and nature of the activities. This book represents a combination of what most instructors agree you need to know and what I think you need to know and believe you want to know.

Some of you may find the contents of this text as exciting as an unexpected promotion (well, almost!), and some of you may find them about as exciting as a Friday afternoon staff meeting. While admittedly some topics are inherently more interesting than others, the important thing is how you approach the content. The stock market report can be engrossing or irrelevant depending upon whether you own any stock. All of

you have some “stock” in evaluation since you all constantly evaluate and are in turn evaluated. Therefore, it is in your best interest to be as skilled and knowledgeable in this area as possible.

It would be next to impossible for you to achieve the objectives of this text and not be a more informed consumer of evaluation information and a more capable conductor of evaluation activities. The basic format of the text goes like this. The task given at the beginning of each part represents a measurement and evaluation activity that you should be able to perform after reading the part. The Performance Criteria at the end of each part give you additional guidance as to how each part is to be done. At the beginning of each chapter a number of subtasks are listed; these activities you should be able to do after reading the chapter. If you can do these subtasks, you shouldn’t have any trouble doing the task. If you are still not sure how the tasks work, read the preface. In fact, it wouldn’t be a bad idea to read the preface anyway (you’ve never read a preface in your whole life, right?!).

Contents

part one: introduction to educational evaluation	3
1 the nature of evaluation	5
Definition and purpose	6
Measurement	8
Evaluation and accountability	10
2 types of evaluation	13
Types versus models	14
Types of evaluation	15
Curriculum evaluation • School evaluation • Evaluation of large populations • Evaluation of special projects and programs • Evaluation of personnel	
3 phases of evaluation	27
The continuity of evaluation	28
The planning phase	29
Situation analysis • Specification of objectives • Specification of prerequisites • Selection and development of measuring instruments • Other procedures	
The process phase	35
The product phase	36
part two: specification of objectives	55

4 the nature of objectives 57

- Definition and purpose 58
 - Definition • Pros and cons
- Needs assessment 63
 - Definition and purpose • Applications • Models

5 classification and statement of objectives 67

- Types of objectives 68
 - Objectives for individuals • Objectives for groups
- The taxonomies of educational objectives 72
 - The cognitive domain • The affective domain • The psychomotor domain
- Levels of specificity 76
 - General objectives—goals • Intermediate objectives • Specific objectives • Degree of specificity
- Stating objectives 81
 - Criteria • Components

part three: classification of tests 101

6 classification of tests 103

- Purpose and process 104
 - Uses of measurement • Data collection • Classification schemes
- Response behaviors 107
 - Written responses • Performance tests
- Data collection methods 118
 - Self-report versus observation • Additional data collection instruments
- Behaviors measured 117
 - Achievement • Character and personality • Aptitude
- Performance standards 129
 - Norm-referenced standards • Criterion-referenced standards • NRTs versus CRTs: summary

part four: selection of a test 153

7 selection of measuring instruments 155

- Characteristics of a test 156
 - Validity • Reliability • Interpretation of validity and reliability coefficients

- Selection of a test 182
 - Selection of measurement methods • Factors to consider •
 - Sources of test information • Selecting from alternatives

part five: development of an instrument 205

8 designing the test 209

- Restriction, definition and selection of test content 210
 - Restriction of content • Definition of content • Selection of test content
- Tables of specifications 212
 - Content outline • Behavioral outcomes • Number of items • A one-way table of specifications

9 constructing test items 217

- Overview 218
- General guidelines 219
 - Objective-item correspondence • Item sufficiency •
 - Communication • Item difficulty • Item novelty • Domain examples
- Essay tests 224
 - Essay item construction • Scoring essay tests
- Objective tests 228
 - Short-answer items • Multiple-choice items • True-false items • Matching items

10 test construction and reproduction 241

- Overview 242
- Organization of items 242
- General directions 243
 - Selection and recording of responses • Scoring information • Correction for guessing
- Format considerations 246
- Test reproduction 247

11 test revision 249

- Rationale 250
- Test validation 250
- Item analysis 251
 - Norm-referenced tests • Criterion-referenced tests • Interpretation of results

part six: analysis and interpretation of results 299

12 analysis 301

- Overview 302
- Types of measurement scales 302
 - Nominal scales • Ordinal scales • Interval scales • Ratio scales
- Descriptive statistics 304
 - Measures of central tendency • Measures of variability • The normal curve • Measures of relative position • Measures of relationship • Calculation for interval data

13 interpretation 331

- Overview 332
- Norm-referenced tests 332
 - Individual profiles • Group profiles
- Criterion-referenced tests 341

part seven: use of results 367

14 using and reporting results 369

- Utilization of results 370
 - Guidance • Instruction
- Grading 373
 - Definition and purpose • General principles of grading • Methods of grading • Types and precision of grading symbols • Alternatives to grading • Promotion
- Reporting procedures 386
 - Reporting within the school system • Reporting to the community

Appendix 1 315

Appendix 2 321

Author Index 433

Subject Index 435

EDUCATIONAL EVALUATION AND MEASUREMENT

The study of evaluation and measurement, as Professor Kingsfield might say, is "something new and unfamiliar to most of you, unlike any other schooling that you have ever known before."



part one

introduction to

educational

evaluation

The study of educational evaluation and measurement, as Professor Kingsfield might say, is "something new and unfamiliar to most of you, unlike any other schooling that you have ever known before." Therefore, in order to meaningfully learn about and perform components of the evaluation process, it is first necessary for you to develop a cognitive structure into which such experiences can be integrated. The purpose of chapters 1-3 is to provide an overview of the evaluation process and of the various types and phases of evaluation. In succeeding chapters specific components of the evaluation/measurement process will be systematically studied and executed.

The goal of part 1 is for you to acquire an understanding of the evaluation/measurement process and related methodology which will facilitate acquisition of specific knowledge and skills. After you have read part 1, you should be able to perform the following task.

task one

Select the type of evaluation of most relevance and interest to you. Briefly describe a specific instance of this type, real or hypothetical, in which you are, will be, or could be involved (e.g., a unit of instruction or a specific program). Briefly describe the factors of interest giving specific examples, which would be involved in the planning phase of evaluation of the above-described endeavor.

See performance criteria, p. 50, and task 1 example, p. 51)

