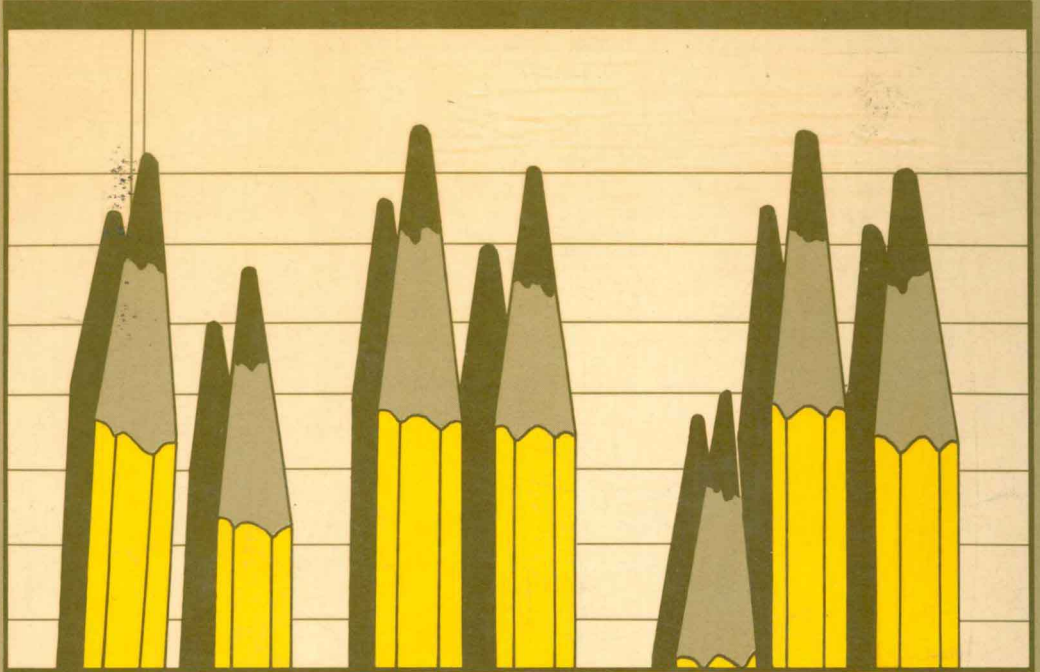


James S. Cangelosi



**DESIGNING TESTS FOR
EVALUATING
STUDENT
ACHIEVEMENT**

DESIGNING TESTS FOR

**EVALUATING
STUDENT
ACHIEVEMENT**

James S. Cangelosi
Utah State University

Designing Tests for Evaluating Student Achievement

Copyright © 1990 by Longman, a division of Addison-Wesley Publishing Co., Inc.
All rights reserved.

No part of this publication may be reproduced,
stored in a retrieval system, or transmitted
in any form or by any means, electronic, mechanical,
photocopying, recording, or otherwise,
without the prior permission of the publisher.

Longman, 95 Church Street, White Plains, N.Y. 10601
A division of Addison-Wesley Publishing Co., Inc.

Associated companies:
Longman Group Ltd., London
Longman Cheshire Pty., Melbourne
Longman Paul Pty., Auckland
Copp Clark Pitman, Toronto

Executive editor: Raymond T. O'Connell
Production editor: Ann P. Kearns
Text design: Jill Francis Wood
Cover design: Anne M. Pompeo
Text art: K & S Graphics, Susan J. Moore, and Martha Bradshaw
Production supervisor: Joanne Jay

Library of Congress Cataloging-in-Publication Data

Cangelosi, James S.

Designing tests for evaluating student achievement / James S. Cangelosi.

p. cm.

Bibliography: p.

Includes index.

ISBN 0-8013-0263-3

✓ 1. Examinations—Design and construction. ✓ 2. Educational tests
and measurements I. Title.

LB3060.65.C36 1990

371.271—dc19

89-30279
CIP

DESIGNING TESTS FOR

**EVALUATING
STUDENT
ACHIEVEMENT**

To Ruth Struyk

Preface

"What are they really learning from this lesson?" "Is Mavis ready to begin working with ratios?" "Should I pick up the pace or slow down this lesson?" "What things motivate these children to read?" "How well did my new strategy for introducing the differences between nouns and pronouns work?" "What grade did Norton earn in history?" "Caroline's father wants a progress report. What should I tell him?" "Should I reteach these sight words one more time?" "Did that last assignment change any attitudes about homework?" "Does Vincent have the psychomotor skills to write legibly?" "Can these students actually apply these principles or have they just memorized them?"

These are just some of the many questions teachers confront every working day. The number and variety of questions about which teachers must make judgments are unparalleled in other professions (Cangelosi, 1974; Clark & Peterson, 1986). To address these questions, teachers must evaluate their students' achievements (i.e., what students have learned or are learning). Thus teachers spend between 20 percent and 30 percent of their time directly involved in data- or information-gathering activities (Stiggins, 1988), including designing, synthesizing, selecting, administering, scoring, interpreting, and revising tests and other types of observations of students' performances and behaviors.

Tests and observations provide the information base for teachers' evaluations of student achievement. Unfortunately, studies examining the validities of tests commonly used in schools (both commercially prepared and teacher-prepared) and the evaluation methods of many teachers suggest that testing malpractice and inaccurate evaluations are widespread (Stiggins, Conklin, & Bridgeford, 1986).

Stiggins (1988, p. 365) points out one of the principal consequences of poorly designed tests:

Teacher-developed paper and pencil tests and many tests and quizzes provided by textbook publishers are currently dominated by questions that ask

students to recall facts and information. Although instructional objectives and even instructional activities may seek to develop thinking skills, classroom assessments often fail to match these aspirations. Students who use tests to try to understand the teacher's expectations can see the priority placed on memorizing, and they respond accordingly. Thus poor quality assessments that fail to tap and reward higher-order thinking skills will inhibit the development of those skills.

There is a brighter side to this story. Many, but probably not most, teachers do manage to collect valid information that leads them to evaluate accurately their students' achievements. They evaluate not only what students have remembered but also how well they conceptualize, comprehend, apply, appreciate, are willing to try, and perform. In other words, teachers are able to evaluate how well students have achieved what is specified by the teachers' learning objectives.

Why do some teachers increase their effectiveness by accurately evaluating student achievement whereas others continue to follow the same outmoded evaluation practices of previous generations? It may be that most teachers have never been exposed to the latest methods for designing achievement tests in a way that is applicable to their needs.

Preservice and in-service teachers who read this book and complete its self-assessment exercises will learn practical ways of applying state-of-the-art strategies for evaluating student achievements. Based on inductive teaching principles for its concept-attainment objectives, deductive teaching principles for its application-level objectives, and expository teaching principles for its knowledge-level objectives, *Designing Tests for Evaluating Student Achievement* includes nine chapters:

1. "Student Achievement" is designed to (a) establish the role of teachers' formative and summative evaluations of student achievement within the overall teaching process, (b) provide a means for teachers to clarify what they mean by "student achievement" and define each learning goal with a set of objectives, and (c) provide a scheme for specifying the content and behavioral construct of each learning objective. As with all other chapters, each new idea or principle is illustrated with real-life, practical classroom examples.
2. "The Measurement of Student Achievement" introduces some ideas, principles, and terms fundamental to the design and selection of quality, cost-effective tests. Topics such as validity and reliability are presented for practical understanding, and esoteric terms and statistical formulas are avoided.
3. "Creating Cost-Effective Achievement Tests" introduces and illustrates a practical seven-step process for constructing valid and usable achievement tests and implementing an achievement-testing management system.
4. "Item Development Hints," the longest chapter, explains a five-step model for designing items and then gives suggestions for building

relevant, reliable, and usable tests with each of the following types of items: short answer, completion, multiple choice, true/false, multiple answer multiple choice, matching, weighted multiple choice, essay, oral discourse, product rating, performance observation, interview, and computer administered.

5. "Items for Cognitive Objectives" suggests how to measure achievement at each of the following cognitive levels: (a) simple knowledge, (b) knowledge of a process, (c) comprehension of a communication, (d) conceptualization, (e) application, and (f) beyond application.
6. "Items for Affective Objectives" suggests how to measure achievement at the following affective levels: (a) appreciation and (b) willingness to act.
7. "Items for Psychomotor Objectives" suggests how to measure achievement at the following psychomotor levels: (a) voluntary muscle capability and (b) ability to perform a specific skill.
8. "Interpreting Standardized Test Scores" illustrates some of the common uses and misuses of standardized tests in schools and explains how to interpret scores (e.g., stanines, percentiles, and grade equivalents) from standardized test reports.
9. "Grading and Reporting Student Achievement" (a) argues for professional behavior by teachers regarding communications of their evaluations of student achievement, (b) describes grade-reporting methods, (c) illustrates and critiques conventional techniques for converting test scores to grades, and (d) introduces and illustrates a new technique for converting scores to grades.

Designing Tests for Evaluating Student Achievement combines ideas from psychometric theory, learning theory, writing and communications models, and commonsense principles discovered through the experiences of classroom teachers.

I owe a debt of gratitude to the hundreds of classroom teachers and students whose ideas and examples influenced the writing of this text. I would like to acknowledge the work of the National Council on Measurement in Education (NCME) for its continuing contributions in the fight to improve the way student achievement is evaluated. NCME publications (e.g., *Journal of Educational Measurement* and *Educational Measurement: Issues and Practice*) emphasize the role teacher-produced tests play in the success of schools.

My dearest friend, Barb Rice, has my thanks for her editing, copyreading, and counsel.

Contents

Preface	xv
Chapter 1 STUDENT ACHIEVEMENT	1
Goal of Chapter 1	1
Difficult Decisions	2
<i>Formative Evaluations</i>	2
<i>Summative Evaluations</i>	2
<i>A Need for Better Evaluation Methods</i>	3
Learning Goals	3
Learning Objectives	6
Content Specified by a Learning Objective	6
Behavioral Construct Specified by a Learning Objective	7
<i>Behavioral Construct Classifications</i>	7
<i>Three Domains</i>	7
<i>The Cognitive Domain</i>	8
<i>Knowledge-Level Cognition</i>	9
<i>Intellectual-Level Cognition</i>	10
<i>The Affective Domain</i>	13
<i>The Psychomotor Domain</i>	13
Using the Scheme for Classifying Behavioral Constructs	15
Self-Assessment of Your Achievement of Chapter 1's Objectives	19
Chapter 2 THE MEASUREMENT OF STUDENT ACHIEVEMENT	23
Goal of Chapter 2	23
Measurements and Evaluations	24
<i>Measurement</i>	25
<i>Evaluation</i>	25

Commonly Used Measurements of Achievement	26
<i>Commercially Produced Tests</i>	26
<i>Standardized Tests</i>	26
<i>Teacher-produced Tests</i>	27
Validity	27
<i>Measurement Relevance</i>	27
<i>Measurement Reliability</i>	29
Assessing Test Validity	32
<i>Examining a Test's Relevance</i>	33
<i>Examining a Test's Reliability</i>	33
<i>Item Analysis</i>	36
Usability	36
Cost-Effectiveness	36
Self-Assessment of Your Achievement of Chapter 2's Objectives	38
 Chapter 3 CREATING COST-EFFECTIVE ACHIEVEMENT TESTS	 41
Goal of Chapter 3	41
A Systematic Approach to Test Construction	42
Clarifying the Learning Goal	42
<i>The First Step in the Process</i>	42
<i>Specifying a Set of Objectives</i>	42
<i>Relative Importance of the Objectives</i>	42
<i>Weighting Objectives</i>	44
<i>Table of Specifications</i>	45
Developing a Test Blueprint	45
<i>The Second Step in the Process</i>	45
<i>The Complexity of the Test Design</i>	45
<i>Administration Time</i>	46
<i>Scoring Time</i>	47
<i>Item Formats</i>	47
<i>Number of Items</i>	47
<i>Difficulty of Items</i>	47
<i>Estimate of the Maximum Number of Points on the Test</i>	47
<i>Number of Points for Each Objective</i>	48
<i>Method for Determining Cutoff Scores</i>	48
<i>Test Outline</i>	49
Obtaining Item Pools	49
<i>The Third Step in the Process</i>	49
<i>The Advantages of Item Pools</i>	49
Desirable Characteristics of an Item Pool	49
<i>Item Components</i>	49
<i>Variety of Item Formats</i>	52
<i>Item Difficulty</i>	53

Maintaining Item Pools	54
<i>File Box System</i>	54
<i>Computerized System</i>	58
Synthesizing the Test	60
<i>The Fourth Step in the Process</i>	60
<i>Item Interaction</i>	60
<i>Sequencing Items</i>	60
<i>Directions to Students</i>	61
Administering the Test	61
<i>The Fifth Step in the Process</i>	61
<i>Minimizing Distractions</i>	61
<i>Following Directions for Test Administration</i>	63
<i>Monitoring Students</i>	63
Preventing Cheating	64
Scoring the Test	69
<i>The Sixth Step in the Process</i>	69
<i>Concern for Usability</i>	69
<i>Concern for Intrascorer and Interscorer Reliabilities</i>	71
Determining Cutoff Scores	71
<i>The Seventh Step in the Process</i>	71
<i>Criterion-referenced Methods</i>	72
<i>Norm-referenced Methods</i>	73
The Most Taxing Task	75
Reviewing Test Results with Students	75
Self-Assessment of Your Achievement of Chapter 3's Objectives	77
 Chapter 4 ITEM DEVELOPMENT HINTS	 79
Goal of Chapter 4	79
Five Critical Steps of Item Design	79
Designing Short-Answer Items	83
<i>Provide-type Objectives</i>	84
<i>Straightforward Directions</i>	85
<i>Danger of Ambiguity</i>	85
<i>Guessing</i>	86
<i>Taxing Communication Skills</i>	86
<i>Structure and Detail</i>	87
<i>Variety of Behavioral Constructs</i>	89
Dichotomously Scored Items	89
Weighted Items	89
<i>Analytically Scored Weighted Items</i>	89
<i>Globally Scored Weighted Items</i>	90
Designing Completion Items	90
<i>Advantage of Structure</i>	90
<i>Too Little Information</i>	91

<i>Too Much Information</i>	91
<i>Frequent Misuse</i>	92
Designing Multiple-Choice Items	93
<i>Directions</i>	93
<i>Design Difficulties</i>	94
<i>Parallel Alternatives</i>	96
<i>Controlling Item Difficulty</i>	96
<i>Correct Responses for the Wrong Reasons</i>	97
<i>Incorrect Responses for the Wrong Reasons</i>	100
<i>Plausible Distractors</i>	101
<i>Placement of the Correct Response</i>	102
<i>Number of Alternatives</i>	102
<i>Correction for Guessing</i>	102
<i>Diagnostic Potential</i>	103
<i>Taxing Comprehension Skills</i>	104
<i>Range of Behavioral Constructs</i>	105
<i>Efficient Use of Words</i>	105
Designing True/False Items	106
<i>The Meaning of "True" and "False"</i>	106
<i>Literally True Statements</i>	107
<i>Ambiguous Statements</i>	109
<i>Descriptive Rather Than Judgmental Statements</i>	110
<i>Number of Items and Guessing</i>	110
<i>Combining True/False and Short-Answer Items</i>	110
Designing Multiple-Answer Multiple-Choice Items	111
Designing Matching Items	113
<i>A Common Basis for the Matches</i>	115
<i>Process of Elimination</i>	115
<i>Brief Responses</i>	115
Designing Weighted Multiple-Choice Items	115
Designing Essay Items	116
<i>Time for Essay Items</i>	117
<i>Flexibility of Expression</i>	117
<i>Taxing Writing Skills</i>	118
<i>Common but Unnecessary Weaknesses</i>	119
<i>Specifying the Task</i>	120
<i>Enhancing Intrascorer and Interscorer Reliabilities</i>	121
Designing Oral Discourse Items	122
Designing Product-rating Items	123
Designing Performance-Observation Items	124
<i>Diagnostic Value</i>	126
<i>Minimizing Observers' Influences on Performances</i>	126
<i>Use of Recording Devices</i>	126
<i>Practical Scoring Forms</i>	126
Designing Interview Items	127
<i>Clarifying</i>	129

<i>Branching and Probing</i>	129
<i>Minimizing the Interviewer's Influence</i>	130
<i>Use of Recording Devices</i>	130
<i>Poor Usability</i>	131
<i>Interviews as Follow-ups to Other Types of Items</i>	131
Computer-administered Tests	131
Self-Assessment of Your Achievement of Chapter 4's Objectives	133
 Chapter 5 ITEMS FOR COGNITIVE OBJECTIVES	 136
Goal of Chapter 5	136
Simple-Knowledge Items	136
<i>Stimulus-Response</i>	136
<i>Avoiding Responses beyond Simple Knowledge</i>	138
Knowledge-of-a-Process Items	139
<i>Stimulus-Response-Response- . . . -Response</i>	139
<i>Emphasis on the Process, Not the Outcome</i>	139
<i>Knowledge, Affective, and Psychomotor Components of the Process</i>	143
Comprehension-of-a-Communication Items	145
<i>Deriving Meanings from Expressions</i>	145
<i>Item Response Mode</i>	145
<i>Novelty</i>	146
Conceptualization Items	148
<i>Concepts and Relationships</i>	148
<i>Grouping Examples</i>	149
<i>Explaining Why</i>	152
Application Items	153
<i>Deciding How to Solve Problems</i>	153
<i>Avoiding "Give-away" Words</i>	154
<i>Mixing Relationships</i>	155
Beyond Application Items	156
Self-Assessment of Your Achievement of Chapter 5's Objectives	157
 Chapter 6 ITEMS FOR AFFECTIVE OBJECTIVES	 160
Goal of Chapter 6	160
Appreciation Items	160
<i>Preferences, Opinions, Desires, and Values</i>	160
<i>Is Appreciation Measurable?</i>	161
<i>Presenting Choices</i>	162
<i>The Direct Approach</i>	162
<i>The Indirect Approach</i>	163
Willingness-to-Act Items	164
<i>Choosing Behaviors</i>	164
<i>Is Willingness to Act Measurable?</i>	165
<i>Observing Behaviors</i>	166

<i>Inferring Behaviors</i>	166
Self-Assessment of Your Achievement of Chapter 6's Objectives	167
Chapter 7 ITEMS FOR PSYCHOMOTOR OBJECTIVES	169
Goal of Chapter 7	169
Voluntary-Muscle-Capability Items	169
<i>Endurance, Strength, Flexibility, Agility, and Speed</i>	169
<i>Isolating the Muscle Group</i>	172
<i>Taxing Knowledge of a Process</i>	173
<i>Fluctuating Physical Conditions</i>	174
<i>Equipment and Environment</i>	174
<i>Repeatable Measurements</i>	174
<i>Caution</i>	174
Ability-to-Perform-a-Specific-Skill Items	175
<i>Execution</i>	175
<i>Focus on Steps in the Process</i>	176
<i>Similarities to Voluntary-Muscle-Capability Items</i>	176
Self-Assessment of Your Achievement of Chapter 7's Objectives	177
Chapter 8 INTERPRETING STANDARDIZED TEST SCORES	178
Goal of Chapter 8	178
District-wide Testing Programs	178
<i>Standardized Tests in Schools</i>	178
<i>A Typical Story</i>	179
<i>Uses and Misuses of Standardized Tests in Schools</i>	183
Standardizing Tests	184
<i>Norm Groups</i>	184
<i>Validation</i>	184
<i>Test Norms</i>	185
Commonly Reported Derived Scores	188
<i>Stanines</i>	188
<i>Percentiles</i>	189
<i>NCE Scores</i>	190
<i>Grade Equivalents</i>	191
<i>Scaled Scores</i>	192
<i>Other Derived Standardized Test Scores</i>	192
Self-Assessment of Your Achievement of Chapter 8's Objectives	192
Chapter 9 GRADING AND REPORTING STUDENT ACHIEVEMENT	196
Goal of Chapter 9	196
Who Should Know about a Student's Achievement	196
<i>Violations of Professional Trust</i>	196

<i>Privileged Information</i>	197
Communicating Evaluations	198
<i>Formative</i>	198
<i>Summative</i>	199
Periodic Reports	199
<i>Clearly Defined Purpose</i>	199
<i>Report Formats</i>	202
<i>Complementing Reports with Parent-Teacher Conferences</i>	202
Determining Test Grades	205
<i>Scoring and Grading</i>	205
<i>Criterion-referenced Grading</i>	205
<i>Norm-referenced Grading</i>	207
<i>The Compromise Method</i>	208
Determining Term Grades	211
<i>The Averaging Method</i>	211
<i>The Weighting Method</i>	211
Self-Assessment of Your Achievement of Chapter 9's Objectives	212
Epilogue TEACHING: A COMPLEX ART	214
All Teachers Evaluate Student Achievement One Way or Another	215
An Evaluation Plan That Works for You	215
Glossary	217
References	223
Index	227

CHAPTER 1

Student Achievement

GOAL OF CHAPTER 1*

Chapter 1 illustrates (1) the vital role played by your evaluations of student achievement and (2) a method for clarifying what you mean by student achievement. More specifically, Chapter 1 will help you to

1. Distinguish between your formative and summative evaluations and clarify the role each plays as you meet your teaching responsibilities (*Cognitive: conceptualization*)
2. Define the learning goal for each unit you teach, using a set of objectives that clarify exactly what students are expected to achieve (*Cognitive: application*)
3. Specify the *content* (i.e., subject matter) of each objective you set for your students (*Cognitive: application*)
4. Specify the *behavioral construct* (i.e., how students are to deal with the content) as either (a) cognitive and simple knowledge, (b) cognitive and knowledge of a process, (c) cognitive and comprehension of a communication, (d) cognitive and conceptualization, (e) cognitive and application, (f) cognitive and beyond application, (g) affective and appreciation, (h) affective and willingness to act, (i) psychomotor and voluntary muscle capability, or (j) psychomotor and ability to perform a specific skill (*Cognitive: application*)
5. State the definition of the following: *formative evaluation, summative evaluation, learning goal, learning objective, content specified by a learning objective, behavioral construct specified by a learning objective, cogni-*

*The goals of this book are defined by objectives appearing at the beginning of each chapter. The terms in parentheses following each objective help clarify its meaning. These terms are defined and explained in Chapter 1 and used throughout the remainder of the book.

tive domain, affective domain, psychomotor domain, simple knowledge, knowledge of a process, comprehension of a communication, conceptualization, application, beyond application, appreciation, willingness to act, voluntary muscle capability, and ability to perform a specific skill (Cognitive: simple knowledge)

According to the classification scheme developed in this chapter, the intended behavioral construct for each of the previously stated objectives is indicated in the parentheses following the objective. You are not expected to understand why these objectives have been so classified until you have completed the chapter.

DIFFICULT DECISIONS

FORMATIVE EVALUATIONS

Consider the decisions confronting Ms. Curry as she conducts a math unit intended to teach her middle school students how to solve surface area problems.

Ms. Curry would like her sixth-graders to extend their understanding of the area of a rectangle to problems about surface areas of other figures, such as triangles. However, she feels that her students will be ready to learn about areas of triangles only after they know how to find rectangular areas and also understand *why* the area of a rectangle is the product of its length and width. Thus, her decision about *when* to teach the area of a triangle depends on her evaluations of how well students can (1) compute rectangular areas and (2) understand *why* a rectangular area equals the length times the width.

When Ms. Curry judges her students' proficiency with rectangular areas in order to decide if they're ready to move on to the next lesson, she is making a *formative evaluation*. *Formative evaluations are judgments about student achievement that influence a teacher's lesson plans.* As a teacher, your continuous evaluations guide what you do next. Should a lesson be extended or terminated? Is remediation needed? Is more advanced work appropriate? Is the pace of the lesson too fast or too slow? Should teaching strategies be altered? Answers to such questions are influenced by feedback from formative evaluations.

SUMMATIVE EVALUATIONS

Consider the decisions Ms. Curry makes at the conclusion of the math unit on surface area problems.