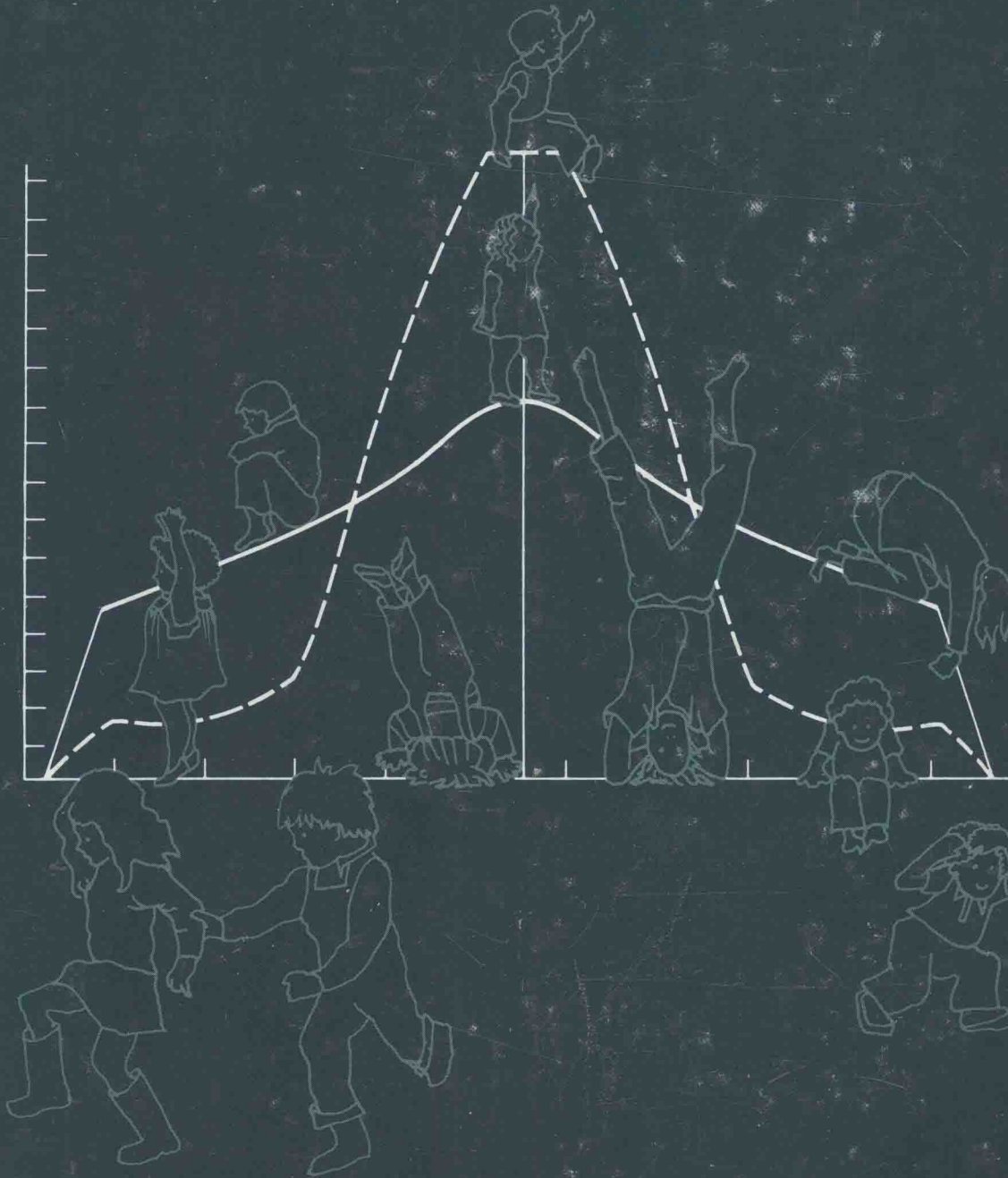


tests, measurement, and evaluation

a developmental approach



arthur bertrand joseph p. cebula

TESTS, MEASUREMENT, AND EVALUATION

A DEVELOPMENTAL APPROACH

ARTHUR BERTRAND and JOSEPH P. CEBULA

American International College



ADDISON-WESLEY PUBLISHING COMPANY

Reading, Massachusetts • Menlo Park, California
London • Amsterdam • Don Mills, Ontario • Sydney

Library of Congress Cataloging in Publication Data

Bertrand, Arthur, 1926-
Tests, measurement & evaluation.

Includes index.

1. Educational tests and measurements.
2. Developmental psychology. I. Cebula, Joseph P., joint author. II. Title.

LB3051.B424 371.2'6

79-21032

ISBN 0-201-00778-9

Copyright © 1980 by Addison-Wesley Publishing Company, Inc. Philippines
copyright 1980 by Addison-Wesley Publishing Company, Inc.

All rights reserved. No part of this publication may be reproduced, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America. Published simultaneously in Canada. Library of Congress Catalog Card No. 79-21032.

TESTS, MEASUREMENT, AND EVALUATION

PREFACE

The testing movement in America has come under severe criticism in recent years by those who claim that schools place too much emphasis on “standardized” instruments to measure intelligence and achievement. The most vocal critics, in fact, have called for a halt to standardized testing because they believe such tests are “dehumanizing” and do not provide an accurate assessment of individual differences. The resultant move away from standardized or “norm-referenced” tests and toward “criterion-referenced” tests has been dramatic.

This book takes the view that tests, in and of themselves, are not dangerous. However, the improper use of tests sometimes makes them seem dangerous. Both norm-referenced and criterion-referenced tests are valuable tools in testing, measuring, and evaluating. Along with systematic observation and teacher-made tests, they provide the classroom teacher with a helpful set of assessment tools.

The emphasis throughout this book is on teachers, whose front-line activities in the day-to-day challenge of the classroom make them a powerful force in the development of their students. And since teachers must interpret test scores directly to parents, they must know how tests are constructed, administered, and interpreted and they must understand as much as possible about the human characteristics they are attempting to measure.

With this in mind, the book takes a developmental approach to growth and learning. It emphasizes the need for teachers to understand each developmental stage of physical, cognitive, and personal growth and how each stage dramatically affects the others throughout a child’s life.

Evaluation, in short, is a dynamic process to be carried on constantly as teachers go about the daily job of assessing the needs of their students. Systematic observation, teacher-made tests, and standardized tests are described in detail as effective ways to help teachers in their assessments and as valuable tools for determining teaching strategies.

The assessment of personal or affective, behavior, is especially emphasized. This aspect of evaluation has often been neglected, perhaps because of the difficulty in assigning quantitative values to "feelings," "attitudes," and "appreciations." Yet these are vitally important facets of growth and development. Good attitudes and feelings about self and life are *taught* and *learned* just as surely as arithmetic. In fact, they play a major role in enhancing or inhibiting the child's success in arithmetic and other "cognitive" subjects.

This book is written for future teachers and teachers already in the field. It supplies a background in testing, test construction, statistics, systematic observation, the writing of objectives, and future trends in developmental psychology. We believe that teachers who understand the nature of the child will be more successful evaluators of the child's physical, cognitive, and personal needs.

We are grateful for the encouragement and help we received in preparing this book. We wish to express special thanks to Dr. Richard Sprinthall for his willingness to read and criticize much of the text and to Dr. Brian Cleary, who contributed his broad knowledge of special education by writing Chapter 11, *Assessing Children in Our Classroom Who Have Special Needs*. We acknowledge our reviewers for the hard but professional look they gave the text and for their helpful suggestions for strengthening our presentation. We acknowledge, with gratitude, the patience, cheerfulness, and expertise of our typists, Brenda Chouinard and Diane Pelletier, and our secretary, Wilhelmina Perkins.

Springfield, Massachusetts
December 1979

A. B.
J. C.

CONTENTS

CHAPTER 1	WHY TEST CHILDREN?	1
	Important points to watch for	1
	Good tests and bad	2
	A test is a tool	3
	The origins of testing	4
	Current criticisms of testing	7
	Teachers test to meet the needs of children	12
	Important points to think about	16
	Notes	17
	Additional readings	17
CHAPTER 2	WHAT DO TEACHERS NEED TO KNOW BEFORE THEY TEST?	18
	Important points to watch for	18
	Teachers need to know about individual differences	19
	A child's behavior is shaped by inheritance and environment	19
	The challenge of evaluating individual needs	20
	The developmental nature of growth	20
	Teachers need to understand physical growth	21
	Physical appearance can affect emotional growth	22
	Teachers need to understand personal growth	23
	Erik Erikson's theories of personal growth	24
	Evaluation of personal growth is difficult but essential	28
	Teachers need to understand intellectual growth	29
	Jean Piaget's theory of cognitive growth	31

	All growth is developmental	36
	Physical, personal, and cognitive growth are closely entwined	37
	Behavior is what teachers test, measure, and evaluate	38
	Important points to think about	39
	Notes	40
CHAPTER 3	THE IMPORTANCE OF OBJECTIVES IN EVALUATING CHILDREN	41
	Important points to watch for	41
	The assessment of learning is not a haphazard process	42
	Objectives help make behavior happen	42
	Three kinds of objectives	43
	Educational goals—the seven cardinal principles	43
	Making objectives more specific	45
	Writing objectives for cognitive and affective learning	49
	Objectives for cognitive behavior	50
	Objectives for affective behavior	52
	Cognitive and affective objectives are both essential to good teaching and evaluation	55
	Bloom's taxonomy of educational objectives	55
	Cognitive and affective evaluation are interrelated	57
	Cognitive and affective objectives for a unit in science	58
	Objectives are vital to the teaching-learning process	60
	Important points to think about	61
	Notes	62
	Additional readings	62
CHAPTER 4	TEACHER ASSESSMENT TOOLS FOR IMPROVING INSTRUCTION: OBSERVING CHILDREN IN THE CLASSROOM	63
	Important points to watch for	63
	Observation of children	64
	Important points to think about	77
	Notes	77
	Additional readings	77

CHAPTER 5	TEACHER ASSESSMENT TOOLS FOR IMPROVING INSTRUCTION: EVALUATING CHILDREN WITH TEACHER-MADE TESTS	78
	Important points to watch for	78
	Using Bloom's taxonomy in preparing tests	79
	Table of specifications	80
	Types of teacher-made tests	84
	Limitations and pitfalls of teacher-made tests	107
	Important points to think about	108
	Notes	109
 CHAPTER 6	 CHARACTERISTICS OF STANDARDIZED TESTS	 110
	Important points to watch for	110
	What are the characteristics of a good test?	110
	Validity	111
	To determine validity, teachers must observe behavior	115
	Tests cannot be valid for all children at all times	116
	Reliability	117
	Testing the test to establish reliability	119
	Correlation	121
	Important points to think about	123
	Additional readings	123
 CHAPTER 7	 STANDARDIZING TESTS AND INTERPRETING RESULTS	 124
	Important points to watch for	124
	What do we mean by standardized?	125
	Statistics needed to interpret test results	126
	Norms	140
	How are standards or norms determined for intelligence tests?	141
	Standardizing the Henmon-Nelson Test of Mental Ability	146
	General intelligence tests are screening tests	150
	Tests of special aptitude	151
	The standardization of achievement tests	153
	Standardizing the Stanford Achievement Test	155
	Determining norms for the test	156
	Administering the test	158

	Summary	158
	Important points to think about	160
	Notes	160
	Additional readings	161
CHAPTER 8	ASSESSING INTELLIGENCE TO IMPROVE INSTRUCTION	162
	Important points to watch for	162
	Intelligence as a hypothetical construct	164
	The language of intelligence	165
	Using intelligence tests to improve instruction	174
	Reviewing the uses of intelligence tests	179
	Important points to think about	185
	Notes	185
CHAPTER 9	ASSESSING ACHIEVEMENT TO IMPROVE INSTRUCTION	187
	Important points to watch for	187
	Achievement and intelligence tests are often similar	188
	We measure achievement by measuring behavior	189
	Standardized achievement tests	190
	Standardized achievement tests are norm-referenced tests	190
	Standardized tests in special areas	191
	Limitations of standardized achievement tests	191
	Informal achievement testing	192
	Informal and teacher-made achievement tests are usually criterion-referenced tests	193
	Criterion-referenced tests and instructional objectives	193
	Three ingredients of criterion-referenced testing	194
	Norm-referenced and criterion-referenced tests are not different kinds of tests	195
	Judgments from norm-referenced test results	195
	Judgments from criterion-referenced test results	196
	How do teachers know test items are appropriate and useful?	197
	Formal criterion-referenced tests	201
	Criterion-referenced tests to meet individual needs	201
	A criterion-referenced reading test	201

	Banks of objectives for criterion-referenced evaluations	204
	When do teachers use norm-referenced tests?	204
	When do teachers use criterion-referenced tests?	205
	Both norm-referenced and criterion-referenced tests are important	206
	Summary	207
	Important points to think about	208
	Notes	208
CHAPTER 10	ASSESSING PERSONAL GROWTH TO IMPROVE INSTRUCTION	209
	Important points to watch for	209
	Good personal growth is essential to learning	209
	The developmental nature of personal and emotional growth	211
	The importance of social relationships to personal growth	212
	Knowing about normal personal growth will help teachers recognize abnormal growth	213
	Assessing the personal strengths of children in the primary grades	213
	Assessing the personal strengths of children in the middle grades	217
	Assessing the personal strengths of children in junior and senior high school	219
	Summary	224
	Important points to think about	225
	Additional readings	226
CHAPTER 11	ASSESSING CHILDREN WITH SPECIAL NEEDS	227
	Important points to watch for	227
	Special education—a changing scene	228
	New laws for the handicapped	228
	Categories of special education	229
	Mental retardation	231
	Emotional disturbance	234
	Children with sensory handicaps	237
	Children with physical handicaps	243
	Children with learning disabilities	245

	Gifted and creative children	249
	Summary	250
	Important points to think about	252
	Notes	252
	Additional readings	252
CHAPTER 12	EVALUATING CHILDREN IN A SCHOOL SYSTEM— A MODEL	254
	Important points to watch for	254
	Testing, measuring, and evaluating children in a small community	255
	Physical development	256
	Personal development	259
	Cognitive development	264
	Special tests of intelligence and aptitude	268
	Reporting developmental growth to others	270
	Summary	272
	Testing, measuring, and evaluating children in the Springmeadow Public Schools	272
	Important points to think about	277
	Additional readings	285
CHAPTER 13	MAJOR ISSUES IN TESTING, MEASURING, AND EVALUATING	286
	Important points to watch for	286
	Teachers need to understand more about testing, measuring, and evaluating	287
	The heritability of intelligence	288
	Accountability for teaching the basics will be expected of everyone	290
	Litigation over testing	292
	Mainstreaming	293
	Culture-fair and culture-free tests	295
	Criterion- versus norm-referenced tests	297
	Summary	298
	Important points to think about	299
	Notes	299
	Additional readings	300

APPENDIX A	GLOSSARY	301
APPENDIX B	AREAS OF THE CURVE	306
APPENDIX C	STATISTICS: SOME REVIEW AND SOME NEW	309
	INDEX	316

WHY TEST CHILDREN?

IMPORTANT POINTS TO WATCH FOR

Tests are tools used to measure and evaluate the growth and development of children.

Tests, in and of themselves, cannot hurt children; professionals who use tests inappropriately can.

The origins of testing and the systematic study of human behavior are very recent. There is much to be learned about both.

Our less than perfect understanding of human behavior and of ways to test it have led to some inappropriate testing practices.

Critics of the testing movement say that tests are dehumanizing, destructive of the ego, and unfair to certain cultural and racial groups. They also claim that there has been too much emphasis placed on the measurement of ability and not enough on the assessment of personal growth.

In the hands of good teachers, tests become valuable tools for the diagnosis of individual strengths and weaknesses.

In the hands of knowledgeable teachers, the *good* things about testing far outweigh the *bad*.

Alice was a fragile and lovely child, painfully shy, almost to the point of withdrawal. Her gentleness was both disturbing and attractive to other children. They were disturbed because she seemed remote from them and from the carefree activities typical of second graders. She seemed older, somehow more mature than they, almost as though she wanted to play but had grown beyond their childlike antics.

They felt attracted to her for almost the same reasons. Her gentle nature seemed to say to them, "I understand how you feel," especially when they were sad or when they had gotten hurt. Without saying a word, she told them by her way of watching that she enjoyed them and liked them. They showed her their art work and their faltering attempts at writing, and she smiled and seemed to say, "That's nice, thank you for showing me." Many of them, with the purity of children, seemed to understand her and make her a part of them.

Alice's teacher understood her, too. At least she thought she did. She saw Alice as a child different from the others, unable to keep up with usual second grade work and unwilling to participate. In her desire to help Alice, she sought to discover whether her shyness and silence were symptoms of some deeper problem. She asked that Alice be tested, and after some months, the school authorities finally gave her an intelligence test. Her I.Q. was found to be 77, a few points below the mandatory cutoff score which, by state law, separated the retarded from the normal. It was recommended that she be placed in a "special class," the designation for classrooms set aside for retarded children.

Alice's teacher, stunned, wondered how this sensitive girl could be retarded, and how a single test could be allowed to uproot her from an atmosphere of relative security and thrust her into a roomful of strangers, many seriously troubled and unable to relate to her with warmth and affection.

Fortunately, her teacher was sensitive enough to children and knowledgeable enough about tests to insist upon additional diagnosis, hypothesizing that, even if the child had problems, retardation wasn't one of them. The results convinced her superiors that emotional distress, and not retardation, caused Alice's poor performance on the I.Q. test. A plan to deal with the reasons for this emotional distress was designed and implemented.

Alice's case is not unusual; its fortunate outcome is. Although she attended school in a northeastern state considered to be a pioneer in education, it was not unusual even in recent years to find children with emotional or behavioral problems labeled "retarded" and removed from the regular classroom. Sometimes it happened because teachers and principals simply could not handle the children in regular classes and sometimes the children were truly retarded. Always there were test scores involved: "Jimmy's I.Q. is only 79; how can he function with 'normal' children?" "Billy has an I.Q. of 80; no wonder he's a discipline problem." These were common statements.

GOOD TESTS AND BAD

Does it seem incredible that in our enlightened society we could make such tragic decisions? Are testing programs bad? and should all testing be

abolished as some critics advocate? We believe not, for in spite of the unfortunate decisions that have been made because of tests, the overwhelming effect of testing has been beneficial. By coming to understand both the positive and negative aspects of testing, you will gain the sensitivity, knowledge, and competency to avoid the potential harm that could occur to students by misuse of tests, and at the same time realize the benefits of the competent use of these tools.

A TEST IS A TOOL

A few important statements need to be made about testing, and you should keep these thoughts in mind as you read this book.

Tests, like thermometers, in and of themselves are not bad. They become bad only in the hands of those who administer and interpret them poorly.

Tests do not hurt children. However, people who use tests can badly hurt children, as you will see in examples throughout this book.

A test is merely a tool, and like any tool, it must be used appropriately to be helpful. A carpenter who uses a chain saw and an axe to build fine cabinets will sell very few cabinets. Similarly a teacher who tests a first grader's arithmetic ability with a calculus exam is not likely to get very promising results.

Measurement, the work for which these tools are designed, is simply one of many ways to evaluate children. In later sections, we will detail the methods used in evaluating children. For now, let's consider a single example.

Alice was evaluated in second grade. The I.Q. test she took yielded a quantitative measurement of her intelligence, 77. Even if the test was an appropriate tool for Alice and the measurement an accurate expression of her ability to solve intellectual problems, it could not provide complete evaluation. Alice's teacher used other tools as well, including careful observation of the child's behavior, especially her ability to share and to engage in activities that require both physical coordination and social skill. She also used tests she made up herself; tools by which she measured Alice's ability to write, read, and speak. A test is only a tool, and it will do its intended job only if the tester knows what it is for and how it is properly used. The measurement that tests provide is only one aspect of the evaluation process.

Testing, measuring, evaluating—these are processes as old as recorded history and among the most criticized and least understood of all teaching tasks. Yet, in spite of the criticisms, can you imagine teaching children multiplication without evaluating their background in arithmetic? Can they count? Can they add and subtract? They must be tested and their arithmetic ability measured before they can learn multiplication. This is the way each new learning experience must begin. Before teachers go to a new

concept or a new skill, they must know whether children have the concepts and skills to build upon. Sometimes the task of measurement is simple, requiring the simple reminder that Timothy is very good at addition and subtraction, while Suzy is less confident and must be asked to do some specific things. By testing her on addition and subtraction, the teacher measures the degree to which she understands these concepts. The evaluation of her work helps her teacher determine whether she is ready for multiplication.

The criticisms leveled against testing are justified to some degree. Educators have been too quick to introduce faulty testing programs into their school systems and too slow to get them out. Test results too often have been used to place children in rigid categories: slow, average, fast, for instance, or retarded, normal, bright. Children have often had great difficulty escaping such categories, not because they couldn't but because those who interpret tests wouldn't allow it. Moreover, the anxiety many children feel when they take a test too often influences the test results. The effects of emotion on testing—fear, apprehension, excitement—have been ignored, and children have been severely penalized for allowing normal feelings to interfere with their ability to answer the questions.

What is it about testing, measuring, and evaluating that has caused such misunderstanding and such criticism? As educators, we need to understand much better than those who have gone before us the limitations of this aspect of teaching. One way is to know the origins of the testing movement and to understand what the first test makers had in mind.

THE ORIGINS OF TESTING

The testing movement as we know it is *less than 100 years old*, a small amount of time in which to build an exact science. The need for testing grew directly out of modern psychology, the study of human behavior which itself originated *less than 100 years ago*. When we consider the centuries over which man has built his theories of astronomy, physics, and mathematics, sciences that continue to change and evolve even today, how can we be so presumptuous as to expect near perfection of testing tools after a mere seventy or eighty years?

In that brief period, an enormous amount of activity ensued as researchers attempted to apply testing procedures to newfound theories of human development. In many instances, the individuals who contributed their efforts to the study of human behavior were the ones who developed methods by which to measure and describe the behavior.

One of the first and most famous of the early pioneers in the testing movement was an English scientist, Sir Francis Galton. At the turn of the nineteenth century, Galton had studied human behavior enough to recognize that individuals differ, not only in physical characteristics, but