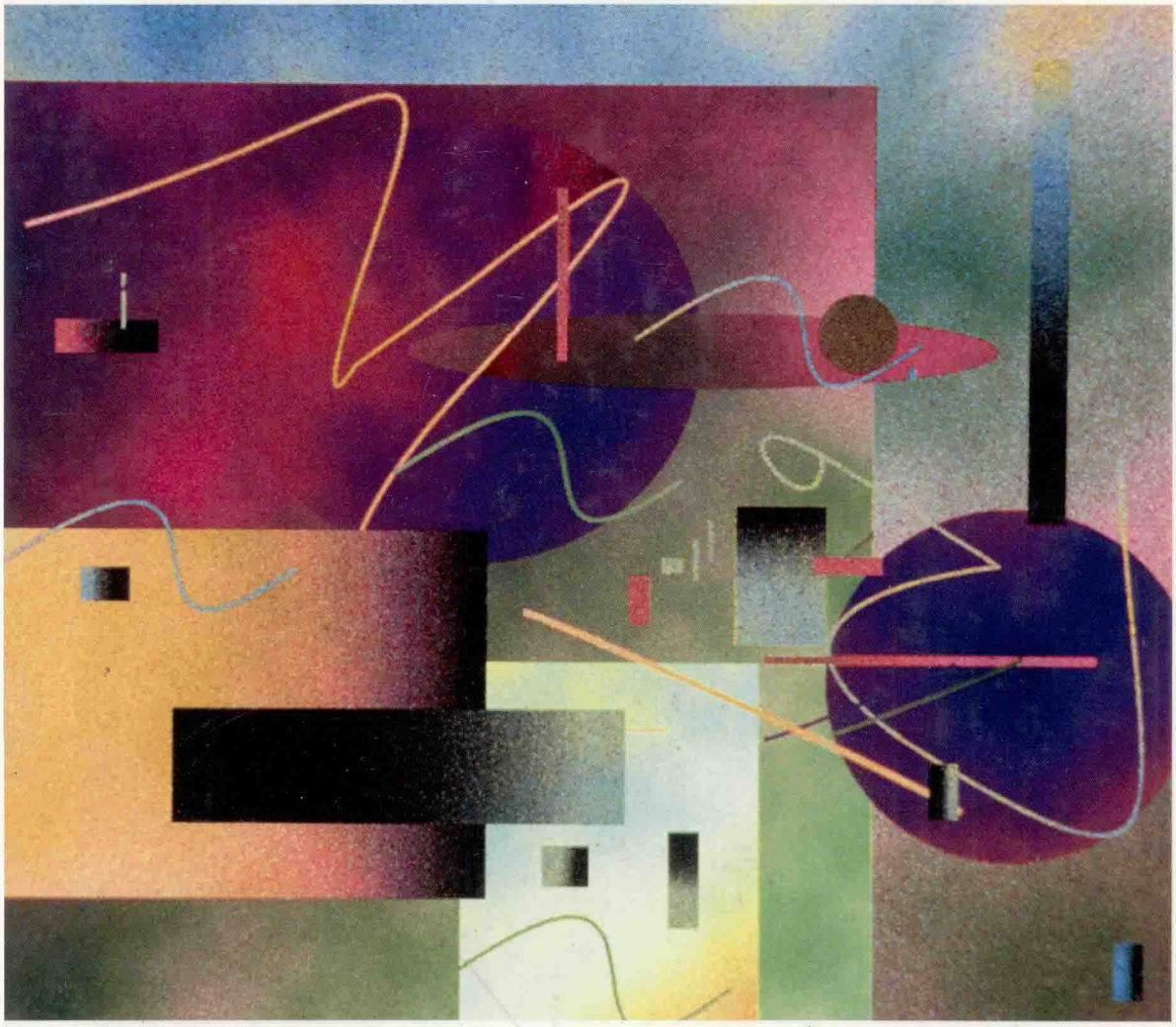
Psychological Testing

AN INTRODUCTION



GEORGE DOMINO

Psychological Testing

An Introduction

GEORGE DOMINO

University of Arizona

Prentice Hall
Upper Saddle River, New Jersey 07458

Library of Congress Cataloging-in-Publication Data

Domino, George
Psychological testing : an introduction / George Domino.
p. cm.
Includes bibliographical references and index.
ISBN 0-13-020143-x (case)
1. Psychological tests. I. Title.
BF176.D65 2000
150'.28'7—dc21 99-19524
CIP

Editor in chief: Nancy Roberts
Executive editor: Bill Webber
Assistant editor: Jennifer Cohen
Editorial/production supervision
and interior design: Mary Araneo
Buyer: Lynn Pearlman
Editorial assistant: Tamsen Adams
Line art coordinator: Guy Ruggiero
Line art creation: Maria Piper
Cover art director: Jayne Conte
Cover designer: Bruce Kenselaar
Cover art: Marjory Dressler

This book was set in 10/11 Times by A & A Publishing Services, Inc.
and was printed and bound by RR Donnelly, Inc. The cover
was printed by Phoenix Color Corp.

© 2000 by Prentice-Hall, Inc.
Upper Saddle River, New Jersey 07458

All rights reserved. No part of this book may be
reproduced, in any form or by any means,
without permission in writing from the publisher.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-13-020143-X

PRENTICE-HALL INTERNATIONAL (UK) LIMITED, *London*
PRENTICE-HALL OF AUSTRALIA PTY. LIMITED, *Sydney*
PRENTICE-HALL CANADA INC., *Toronto*
PRENTICE-HALL HISPANOAMERICANA, S.A., *Mexico*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*
PEARSON EDUCATION ASIA PTE. LTD., *Singapore*
EDITORA PRENTICE-HALL DO BRASIL, LTDA., *Rio de Janeiro*

Preface

My first professional publication, in 1963, was as a graduate student (with Harrison Gough) on a validational study of a culture-fair test. Since then, I have taught a course on psychological testing with fair regularity. At the same time I have steadfastly refused to specialize and have had the opportunity to publish in several different areas, to work in management consulting, to be director of a counseling center and a clinical psychology program, to establish an undergraduate honors program, and to be involved in a wide variety of projects with students in nursing, rehabilitation, education, social work, and other fields. In all of these activities I have found psychological testing to be central and to be very challenging and exciting.

In this book I have tried to convey the excitement associated with psychological testing and to teach basic principles through the use of concrete examples. When specific tests are mentioned, they are mentioned because they are used as an example to teach important basic principles, or in some instances because they occupy a central/historical position. No attempt has been made to be exhaustive.

Much of what is contained in many textbooks is rather esoteric information of use only to very few readers. For example, most textbooks include several formulas to compute interitem consistency. However, it has been my experience that 99% of the students who take a course on testing will never have occasion to use such formulas, even if they enter a career in psychology or allied fields. The very few who might need to do such calculations will do them by computer or will know where to find the relevant formulas. It is the principle that is important, and that is what I have tried to emphasize.

Because of my varied experience in industry, in a counseling center and other service oriented settings, and also because as a clinically trained academic psychologist I have done a considerable amount of research, I have tried to cover both sides of the coin—the basic research-oriented issues and the application of tests in service-oriented settings. Thus parts I and II, the first eight chapters, serve as an introduction to basic concepts, issues, and approaches. Parts III and IV, Chapters 9 through 15, have a much more applied focus. Finally, I have attempted to integrate both classical approaches and newer thinking about psychological testing.

The area of psychological testing is fairly well defined. I cannot imagine a textbook that does not discuss such topics as reliability, validity, norms, etc. Thus what distinguishes one textbook from another is not so much their content as a question of balance. For example, most textbooks continue to devote one or more chapters to projective techniques, even though their use and importance has decreased substantially. Projective techniques are important, not only from a historical perspective,

but also for what they can teach us about basic issues in testing. In this text, they are discussed and illustrated, but as part of a chapter (see Chapter 15) within the broader context of testing in clinical settings. Most textbooks also have several chapters on intelligence testing, often devoting considerable space to such topics as the heritability of intelligence, theories of trait organization, longitudinal studies of intelligence, and similar topics. Such topics are of course important and fascinating, but do they really belong in a textbook on psychological testing? If they do, then that means that some other topics more directly relevant to testing are omitted or given short shrift. In this textbook, I have chosen to focus on testing, and to minimize the theoretical issues associated with intelligence, personality, etc., except where they may be needed to have a better understanding of testing approaches.

It is no surprise that computers have had (and continue to have) a major impact on psychological testing, and so an entire chapter (Chapter 17) is devoted to this topic. There is also a vast literature and great student interest on the topic of faking, and here too an entire chapter (Chapter 16) has been devoted to this topic. Most textbooks begin with a historical chapter. I have chosen to place the history chapter as last so that having learned about testing the reader can now learn about the historical background from a more knowledgeable point of view.

Finally, rather than writing a textbook about testing, I have attempted to write a textbook about testing the individual. That is, I believe that most testing applications involve an attempt to use tests as a tool in order to better understand an individual, whether that person is a client in therapy, a college student seeking career or academic guidance, a business executive wishing to capitalize on strengths and improve weaknesses, or a subject in a scientific experiment.

ACKNOWLEDGMENTS

In my career as a psychologist, I have had the excellent fortune of being mentored, directly and indirectly, by three giants in the psychological testing field. The first is Harrison Gough, my mentor in graduate school at Berkeley, who showed me how useful and exciting psychological tests can be when applied to real life problems. More important, Gough has continued to be not only a mentor but a genuine model to be emulated both as a psychologist and as a human being. Much of my thinking, my approach to testing, and my major interest in students at all levels is a direct reflection of Gough's teaching.

The second is Anne Anastasi, who was a treasured colleague at Fordham University, a generous friend, and one of the best chairpersons I have ever worked with. Her textbook has been truly a model of scholarship and concise writing, the product of an extremely keen mind who has advanced the field of psychological testing in many diverse ways.

The third person is Lee J. Cronbach of Stanford University. My first undergraduate exposure to testing was through his textbook. In 1975 Cronbach wrote what is now a classic paper titled "Beyond the two disciplines of scientific psychology" (*American Psychologist,* 1975, volume 30, pp. 116–127), in which he argued that experimental psychology and the study of individual differences should be integrated. In that paper, Cronbach was kind enough to cite at some length two of my studies on college success as examples of this integration. Subsequently I was able to invite him to give a colloquium at the University of Arizona. My contacts with him have been regrettably brief, but his writings have greatly influenced my own thinking.

Additionally, I would like to thank the following colleagues who read, reviewed, and commented on the manuscript: James F. Austin of the University of Akron, Sylvia Marotta of George Washington University, and Peggy Parsons of Xavier University

On a personal note, I must first and foremost thank Valerie, my wife of 34 years, for her love and support, and for being the best companion one could hope for in this voyage we call life. Secondly, our three children have been an enormous source of love and pride: Brian, who is currently a professor of philosophy at Eastern Michigan University; Marisa, a postdoctoral fellow in health economics at Harvard; and Marla, a

此为试读,需要完整PDF请访问: www.ertongbook.com

doctoral candidate in clinical psychology at the University of Alabama. I have also been truly blessed by some exceptional friends, whose love and caring have been truly supportive in my efforts: George and Betty, Ray and Jeanne, Clive and Mary Jean, John and Linda, and Marv and Linda. They have enriched my life to an unbelievable degree.

As a faculty member I have had the exciting opportunity to relate to hundreds of students over the years. I hope I have been a positive influence on most of them, but I could easily mention dozens of names of students who in one way or another greatly enriched my intellectual and personal life. In particular, a sincere thanks to Dyanne Affonso, Gary Blair, Karen Miller, and LucyJo Palladino.

George Domino

Contents

Preface ix

PART ONE: BASIC ISSUES

Chapter 1
The Nature of Tests 1
Aim 1
Introduction 1
Categories of Tests 5
Ethical Standards 10
Information about Tests 12
Summary 13
Suggested Reading 15

Chapter 2
Test Construction, Administration, and Interpretation 17
Aim 17
Constructing a Test 17
Test Items 20
Philosophical Issues 24
Administering a Test 27
Interpreting Test Scores 27
Item Characteristics 31
Norms 36
Combining Test Scores 39
Summary 42
Suggested Reading 42

Chapter 3
Reliability and Validity 44
Aim 44
Introduction 44
Reliability 44
Types of Reliability 45
Validity 54
Aspects of Validity 59
Summary 67
Suggested Reading 68

PART TWO: DIMENSIONS OF TESTING

Chapter 4
Personality 69
Aim 69
Introduction 69
Some Basic Issues 70
Types of Personality Tests 72
Examples of Specific Tests 74
The Big Five 93
Summary 95
Suggested Reading 96

Chapter 5
Cognition 97
Aim 97
Introduction 97
Theories of Intelligence 97
Other Aspects 102
The Binet Tests 105
The Wechsler Tests 110
Other Tests 121
Summary 130
Suggested Reading 131

Chapter 6
Attitudes, Values, and Interests 132
Aim 132
Attitudes 132
Values 146
Interests 156
Summary 167
Suggested Reading 168

Chapter 7
Psychopathology 169
Aim 169
Introduction 169
Measures 172
The Minnesota Multiphasic Personility Inventory (MMPI) and MMPI-2 178
The Millon Clinical Multiaxial Inventory (MCMI) 187
Other Measures 194
Summary 205
Suggested Reading 205

Chapter 8
Normal Positive Functioning 206
Aim 206
Self-concept 206
Locus of Control 212
Sexuality 214
Creativity 215
Imagery 223
Competitiveness 225
Hope 226
Hassles 228
Loneliness 228
Automobile Driving 229
Death Anxiety 230
Summary 231
Suggested Reading 231

PART THREE: APPLICATIONS OF TESTING

Chapter 9
Special Children 232
Aim 232
Some Issues Regarding Testing 232
Categories of Special Children 243
Some General Issues about Tests 256
Summary 266
Suggested Reading 266

Chapter 10
Older Persons 268
Aim 268
Some Overall Issues 268
Attitudes Toward the Elderly 271
Anxiety about Aging 272
Life Satisfaction 272
Marital Satisfaction 274
Morale 275
Coping or Adaptation 276
Death and Dying 277
Neuropsychological Assessment 277
Depression 281
Summary 282
Suggested Reading 282

Chapter 11
Testing in a Cross-Cultural Context 284
Aim 284
Introduction 284
Measurement Bias 284
Cross-Cultural Assessment 294
Measurement of Acculturation 297
Some Culture-Fair Tests and Findings 300
Standardized Tests
Summary 308
Suggested Reading 308

Chapter 12
Disability and Rehabilitation 310
Aim 310
Some General Concerns 310
Modified Testing 313
Some General Results 314
Legal Issues 317
The Visually Impaired 321
Hearing Impaired 325
Physical-Motor Disabilities 335
Summary 338
Suggested Reading 338

PART FOUR: THE SETTINGS

Chapter 13
Testing in the Schools 339
Aim 339
Preschool Assessment 339
Assessment in the Primary Grades 342
High School 345
Admission into College 349

The Graduate Record Examination 356
Entrance into Professional Training 363
Tests for Licensure and Certification 367
Summary 369
Suggested Reading

Chapter 14
Occupational Settings 371
Aim 371
Some Basic Issues 371
Some Basic Findings 374
Ratings 375
The Role of Personality 379
Biographical Data (Biodata) 381
Assessment Centers 386
Illustrative Industrial Concerns 388
Testing in the Military 391
Prediction of Police Performance 393
Examples of Specific Tests 395
Integrity Tests 400
Summary 404
Suggested Reading 404

Chapter 15
Clinical and Forensic Settings 405
Aim 405
Clinical Psychology: Neuropsychological Testing 405
Projective Techniques 407
Some Clinical Issues and Syndromes 422
Health Psychology 426
Forensic Psychology 436
Legal Cases 439
Summary 442
Suggested Reading 442

PART FIVE: CHALLENGES TO TESTING

Chapter 16
The Issue of Faking 444
Aim 444
Some Basic Issues 444
Some Psychometric Issues 449
Techniques to Discourage Faking 451
Related Issues 453
The MMPI and Faking 455
The CPI and Faking 460
Social Desirability and Assessment Issues 462
Acquiescence 465
Other Issues 467
Test Anxiety 474
Testwiseness 475
Summary 476
Suggested Reading 476

Chapter 17
The Role of Computers 478
Aim 478
Historical Perspective 478
Computer Scoring of Tests 479
Computer Administration of Tests 480
Computer-Based Test Interpretations (CBTI) 485
Some Specific Tests 489
Adaptive Testing and Computers 491
Ethical Issues Involving Computer Use 494
Other Issues and Computer Use 495
A Look at Other Tests and Computer Use 496
The Future of Computerized Psychological Testing 499
Summary 500
Suggested Reading 500

Chapter 18
Testing Behavior and Environments 501
Aim 501
Traditional Assessment 501
Behavioral Assessment 502
Traditional vs. Behavioral Assessment 506
Validity of Behavioral Assessment 506
Behavioral Checklists 509
Behavioral Questionnaires 510
Program Evaluation 521
Assessment of Environments 522
Assessment of Family Functioning 526
Broad-Based Instruments 530
Summary 535
Suggested Reading 535

Chapter 19
The History of Psychological Testing 537
Aim 537
Introduction 537
French Clinical Tradition 538
German Nomothetic Approach 539
British Idiographic Approach 541
American Applied Orientation 542

Some Recent Developments 550
Summary 553
Suggested Reading 553

Appendix
Table to Translate Difficulty Level of a Test Item into a *z* Score 555

References 557

Test Index 642

Index of Acronyms 646

Subject Index 648

Chapter 1

The Nature of Tests

AIM

In this chapter we will cover four basic issues. The first centers on what is a test, not just a formal definition, but ways of thinking about tests. Second, we will try to develop a "taxonomy" of tests, that is we will look at various ways in which tests can be categorized. Third, we will look at the ethical aspects of psychological testing. Finally, we will explore how we can obtain information about a specific test.

INTRODUCTION

Most likely you would have no difficulty identifying a psychological test, even if you met one in a dark alley. So the intent here is not to give you one more definition to memorize and repeat but rather to spark your thinking.

What is a test? Anastasi (1988), one of the best known psychologists in the field of testing, defined a test as an "objective" and "standardized" measure of a sample of behavior. This is an excellent definition that focuses our attention on *three* elements: (1) *objectivity*: that is, at least theoretically, most aspects of a test, such as how the test is scored and how the score is interpreted, are not a function of the subjective decision of a particular examiner but are based on objective criteria; (2) *standardization*: that is, no matter who administers, scores, and interprets the test, there is uniformity of procedure; and (3) *a sample of behavior*: a test is not a psychological X-ray, nor does it necessarily reveal hidden conflicts and forbidden wishes; it is a sample of a person's behavior, hopefully a representative sample from which we can draw some inferences and hypotheses.

There are three other ways to consider psychological tests which I find useful and I hope you will also. One way is to consider the administration of a test as an experiment. In the classical type of experiment, the experimenter studies a phenomenon and observes the results, while at the same time keeping in check all extraneous variables so that the results can be ascribed to a particular antecedent cause. In psychological testing, however, it is usually not possible to control all the extraneous variables, but the metaphor here is a useful one that forces us to focus on the standardized procedures, on the elimination of conflicting causes, on experimental control, and on the generation of hypotheses that can be further investigated. So if I administer a test of achievement to little Sandra, I want to make sure that her score reflects what she has achieved, rather than her ability to follow instructions, her degree of hunger before lunch, her uneasiness at being tested, or some other influence.

A second way to consider a test is to think of a test as an interview. When you are administered an examination in your class, you are essentially being interviewed by the instructor to determine how well you know the material. We will discuss

此为试读,需要完整PDF请访问: www.ertongbook.com

interviews in Chapter 18, but for now consider the following: in most situations we need to "talk" to each other. If I am the instructor I need to know how much you have learned. If I am hiring an architect to design a house or a contractor to build one, I need to evaluate their competency, and so on. Thus "interviews" are necessary, but a test offers many advantages over the standard interview. With a test I can "interview" 50 or 5000 persons at one sitting. With a test I can be much more objective in my evaluation, since for example, multiple-choice answer sheets do not discriminate on the basis of gender, ethnicity, or religion.

A third way to consider tests is as tools. Many fields of endeavor have specific tools—for example, physicians have scalpels and X-rays, chemists have Bunsen burners and retorts. Just because someone can wield a scalpel or light up a Bunsen burner does not make him or her an "expert" in that field. The best use of a tool is in the hands of a trained professional when it is simply an aid to achieve a particular goal. Tests, however, are not just psychological tools; they also have political and social repercussions. For example, the well-publicized decline in SAT scores (Wirtz & Howe, 1977) has been used as an indicator of the terrible shape our educational system is in (National Commission, 1983).

A test by any other name. . . . In this book we will use the term *psychological test* (or more briefly *test*) to cover those measuring devices, techniques, procedures, examinations, etc., that in some way assess variables relevant to psychological functioning. Some of these variables, such as intelligence, introversion- extraversion, and self-esteem are clearly "psychological" in nature. Others, like heart rate or the amount of palmar perspiration (the galvanic skin response), are more physiological but are related to psychological functioning. Still other variables, like socialization, delinquency, or leadership may be somewhat more "sociological" in nature, but are of substantial interest to most social and behavioral scientists. Other variables, like academic achievement, might be more relevant to educators or professionals working in educational settings. The point here is that we use the term *psychological* in a rather broad sense.

Psychological tests can take a variety of forms. Some are true-false *inventories*, others are *rating scales*, some are actual *tests* whereas others are *questionnaires*. Some tests consist of materials like inkblots or pictures to which the subject responds verbally; still others consist of items like blocks or pieces of a puzzle which the subject manipulates. A large number of tests are simply a set of printed items requiring some type of written response.

Testing versus assessment. *Psychological assessment* is basically a judgmental process whereby a broad range of information, often including the results of psychological tests, is integrated into a meaningful understanding of a particular person. If that person is a client or patient in a psychotherapeutic setting, we call the process *clinical assessment*. Psychological testing is thus a narrower concept referring to the psychometric aspects of a test (the technical information about the test), the actual administration and scoring of the test, and the interpretation made of the scores. We could of course assess a client simply by administering a test or *battery* (group) of tests. Usually the assessing psychologist also interviews the client, obtains background information, and where appropriate and feasible, information from others about the client (see Korchin, 1976, for an excellent discussion of clinical assessment).

Who uses tests? Oakland and Hu (1991) asked some 53 test experts in 44 different countries "Who uses tests?" They reported that some 16 professional groups used tests, including school psychologists, special-education teachers, clinical psychologists, guidance counselors, psychiatrists, speech therapists, guidance teachers, nurses, and even engineers. It is interesting to note that among the 44 countries surveyed, 30 used more foreign than locally developed tests.

Purposes of tests. Tests are used for a wide variety of purposes that can be subsumed under more general categories. Many authors identify four categories typically labelled as *classification*, *self-understanding*, *program evaluation*, and *scientific inquiry*.

Classification involves a decision that a particular person belongs in a certain category. For

example, based on test results we may assign a diagnosis to a patient, accept or reject a college applicant, place a student in the introductory Spanish course rather than the intermediate or advanced course, or certify that a person has met the minimal qualifications to practice medicine. A crucial distinction is that of diagnosis versus mastery; we'll see later, these types of tests have somewhat different properties.

Self-understanding involves using test information as a source of information about oneself. Such information may already be available to the individual, but not in a formal way. Marlene, for example, is applying to graduate studies in electrical engineering; her high GRE scores confirm what she already knows, that she has the potential abilities required for such study.

Program evaluation involves the use of tests to assess not so much the individual testee, but the effectiveness of a particular program or course of action. You have probably seen in the newspaper tables indicating the average achievement test scores for various schools in your geographical area, with the scores often taken, perhaps incorrectly, as evidence of the competency level of a particular school. Program evaluation may involve the assessment of the campus climate at a particular college, or the value of a drug abuse program offered by a mental health clinic, or the effectiveness of a new medication.

Tests are also used in scientific inquiry. If you glance through most professional journals in the social and behavioral sciences, you will find that a large majority of studies use psychological tests to operationally define relevant variables and to translate hypotheses into numerical statements that can be assessed statistically. Some argue that development of a field of science is in large part a function of the available measurement techniques (Cone & Foster, 1991; Meehl, 1978).

Tests as experimental procedure. If we accept the analogy that administering a test is very much like an experiment, then we need to make sure that the experimental procedure is followed carefully and that extraneous variables are not allowed to influence the results. This means, for example, that instructions and time limits need to be adhered to strictly. The greater the control that can be exercised on all aspects of a test situation, the lesser the influence of extraneous variables. Thus the scoring of a multiple-choice exam is less influenced by such variables as clarity of handwriting than the scoring of an essay exam; a true-false personality inventory with simple instructions is probably less influenced than an intelligence test with detailed instructions.

Masling (1960) reviewed a variety of studies of variables that can influence a testing situation, in this case "projective" testing (see Chapter 15); Sattler and Theye (1967) did the same for intelligence tests. We can identify, as Masling (1960) did, four categories of such variables:

1. *The method of administration.* Standard administration can be altered by disregarding or changing instructions, by explicitly or implicitly giving the subject a set to answer in a certain way, or by not following standard procedures. For example, Coffin (1941) had subjects read fictitious magazine articles indicating what were more socially acceptable responses to the Rorschach Inkblot test. Subsequently they were tested with the Rorschach and the responses clearly showed a suggestive influence on the part of the prior readings. Ironson and Davis (1979) administered a test of creativity three times, with instructions to "fake creative," "fake uncreative," or "be honest"; the obtained scores reflected the influence of the instructions. On the other hand, Sattler and Theye (1967) indicated that of twelve studies reviewed that departed from standard administrative procedures, only five reported significant differences between standard and nonstandard administration.

2. *Situational variables.* These include a variety of aspects that presumably can alter the test situation significantly, such as a subject feeling frustrated, discouraged, hungry, being under the influence of drugs, and so on. Some of these variables can have significant effects on test scores, but the effects are not necessarily the same for all subjects. For example, Sattler and Theye (1967) report that discouragement affects the performance of children but not of college students on some intelligence tests.

3. *Experimenter variables.* The testing situation is a social situation, and even when the test is administered by computer, there is clearly an experimenter, a person in charge. That person

may exhibit characteristics (such as age, gender, and skin color) that differ from those of the subject. The person may appear more or less sympathetic, warm or cold, more or less authoritarian, aloof, more adept at establishing rapport, etc. These aspects may or may not affect the subject's test performance; the results of the available experimental evidence are quite complex and not easily summarized. We can agree with Sattler and Theye (1967) who concluded that the experimenter-subject relationship is important and that (perhaps) less qualified experimenters do not obtain appreciably different results than more qualified experimenters. Whether the race, ethnicity, physical characteristics, etc., of the experimenter significantly affect the testing situation seems to depend on a lot of other variables and, in general, does not seem to be as powerful an influence as many might think.

4. *Subject variables.* Do aspects of the subject, such as level of anxiety, physical attractiveness, etc., affect the testing situation? Masling (1960) used attractive female accomplices who, as test subjects, acted "warm" or "cold" toward the examiners (graduate students). The test results were interpreted by the graduate students more favorably when the subject acted warm than when she acted cold.

In general what can we conclude? Aside from the fact that most studies in this area seem to have major design flaws and that many specific variables have not been explored consistently, Masling (1960) concluded that there is strong evidence of situational and interpersonal influences in projective testing, while Sattler and Theye (1967) concluded that:

1. Departures from standard procedures are more likely to affect "specialized" groups, such as children, schizophrenics, and juvenile delinquents than "normal" groups such as college students;
2. Children seem to be more susceptible to situational factors, especially discouragement, than are college-age adults;
3. Rapport seems to be a crucial variable, while degree of experience of the examiner is not;
4. Racial differences, specifically a white examiner and a black subject, may be important, but the evidence is not definitive.

Tests in decision making. In the real world decisions need to be made. To allow every person who applies to medical school to be admitted would not only create huge logistical problems, but would result in chaos and in a situation that would be unfair to the candidates themselves, some of whom would not have the intellectual and other competencies required to be physicians, to the medical school faculty whose teaching efforts would be diluted by the presence of unqualified candidates, and eventually to the public who might be faced with incompetent physicians.

Given that decisions need to be made, we must ask what role psychological tests can play in such decision making. Most psychologists agree that major decisions should not be based on the results of a single test administration, that whether or not state university admits Sandra should not be based solely on her SAT scores. In fact, despite a stereotype to the contrary, it is rare for such decisions to be based solely on test data. Yet in many situations, test data represents the only source of objective data that is standard for all candidates; other sources of data such as interviews, grades, and letters of recommendation are all "variable" —grades from different schools or different instructors are not comparable, nor are letters written by different evaluators. Finally, as scientists, we should ask what is the empirical evidence for the accuracy of predicting future behavior. That is, if we are admitting college students to a particular institution, which sources of data, singly or in combination, such as interviewers' opinions, test scores, high school GPA, etc., would be most accurate in making relevant predictions, such as "let's admit Marlene since she will do quite well academically." We will return to this issue, but for now let me indicate a general psychological principle that past behavior is the best predictor of future behavior, and a corollary that the results of psychological tests can provide very useful information on which to make more accurate future predictions.

Relation of test content to predicted behavior. Rebecca is enrolled in an introductory Spanish

course and is given a Spanish vocabulary test by the instructor. Is the instructor interested in whether Rebecca knows the meaning of the specific words on the test? Yes, indeed, since the test is designed to assess Rebecca's mastery of the vocabulary that has been covered in class and in the homework assignments. Consider now a test like the SAT, given for college admission purposes. The test may contain a vocabulary section, but the concern is not whether an individual knows the particular words; knowledge of this sample of words is related to something else, namely doing well academically in college. Finally consider a third test, the XYZ scale of depression. Although the scale contains no items about suicide ideation, it has been discovered empirically that high scorers on this scale are likely to attempt suicide. These three examples illustrate an important point: in psychological tests, the content of the test items may or may not cover the behavior that is of interest—there may be a lack of correspondence between test items and the predicted behavior. But a test can be quite useful if an empirical correspondence between test scores and real life behavior can be shown.

CATEGORIES OF TESTS

Since there are thousands of tests, it would be helpful to be able to classify tests into categories, just as a bookstore might list its books under different headings. Because tests differ from each other in a variety of ways there is no uniformly accepted system of classification. Therefore we will invent our own based on a series of questions that can be asked of any test. I should point out that despite a variety of advances in both theory and technique, standardized tests have changed relatively little over the years (Linn, 1986), so while new tests are continually published, a classificatory system should be fairly stable, i.e., applicable today as well as 20 years from now.

Commercially published? The first question is whether a test is commercially published (sometimes called a proprietary test) or not. Major tests like the Stanford-Binet and the Minnesota Multiphasic Personality Inventory are available for purchase by qualified users through commercial companies. The commercial publisher advertises primarily through its catalog, and for many tests makes available, for a fee, a *specimen set*, usually the test booklet and answer sheet, a scoring key to score the test, and a test manual that contains information about the test. If a test is not commercially published, then a copy is ordinarily available from the test author, and there may be some accompanying information, or perhaps just the journal article where the test was first introduced. Sometimes journal articles will include the original test, particularly if it is quite short, but often they will not. (Examples of articles which contain the test items are Baker, Mednick & Hocevar, 1991; Good & Good, 1974; McLain, 1993; Rehfisch, 1958a; Snell, 1989; Vodanovich & Kass, 1990). Keep in mind that the contents of journal articles are copyrighted and permission to use a test must be obtained from both the author and the publisher.

If you are interested in learning more about a specific test, first you must determine if the test is commercially published. If it is, then you will want to consult the *Mental Measurements Yearbook* (MMY), available in most university libraries. Despite its name, the MMY is published at irregular intervals rather than yearly. However, it is an invaluable guide. For many commercially published tests the MMY will provide a brief description of the test (its purpose, applicable age range, type of score generated, price, administration time, and name and address of publisher), a bibliography of citations relevant to the test, and one or more reviews of the test by test experts. Tests that are reviewed in one edition of the MMY may or may not be reviewed in subsequent editions, so locating information about a specific test may involve browsing through a number of editions. MMY reviews of specific tests are also available through a computer service called the Bibliographic Retrieval Services.

If the test you are interested in learning about is not commercially published, it will probably have an author(s) who published an article about the test in a professional journal. The journal article will most likely give the author's address at the time of publication. If you are a "legitimate" test user, for example a graduate student doing a

doctoral dissertation or a psychologist engaged in research work, a letter to the author will usually result in a reply with a copy of the test and permission to use it. If the author has moved from the original address you may locate the current address through various directories and "Who's Who" type of books, or through computer-generated literature searches.

Administrative aspects. Tests can also be distinguished by various aspects of their administration. For example, there are *group* versus *individual* tests; group tests can be administered to a group of subjects at the same time and individual tests to one person only at one time. The Stanford-Binet test of intelligence is an individual test, whereas the SAT is a group test. Clinicians who deal with one client at a time generally prefer individual tests since these often yield observational data in addition to a test score; researchers often need to test large groups of subjects in minimum time and may prefer group tests (there are of course, many exceptions to this statement). A group test can be administered to one individual; sometimes, an individual test can be modified so it can be administered to a group.

Tests can also be classified as *speed* versus *power* tests. Speed tests have a time limit which affects performance; for example, you might be given a page of printed text and asked to cross out all the "e's" in 25 seconds. How many you cross out will be a function of how fast you respond. A power test on the other hand is designed to measure how well you can do; and so either may have no time limit or a time limit of convenience (a 50-minute hour) that ordinarily does not affect performance. The time limits on speed tests are usually set so that only 50% of the applicants are able to attempt every item. Time limits on power tests are set so that about 90% of the applicants can attempt all items.

Another administrative distinction is whether a test is a *secure* test or not. For example, the SAT is commercially published but is ordinarily not made available even to researchers. Many tests that are used in industry for personnel selection are secure tests whose utility could be compromised if they were made public. Sometimes only the scoring key is confidential, rather than the items themselves.

A final distinction from an administrative point of view is how *invasive* the test is. A questionnaire that asks about one's sexual behaviors is ordinarily more invasive than a test of arithmetic; a test completed by the subject is usually more invasive than a report of an observer, who may report the observations without even the subject's awareness.

The medium. Tests differ widely in the materials used, and so we can distinguish tests on this basis. Probably, the majority of tests are *paper-and-pencil* tests that involve some set of printed questions and require a written response, like marking a multiple answer sheet. Other tests are *performance* tests that perhaps require the manipulation of wooden blocks or the placement of puzzle pieces in correct juxtaposition. Still other tests involve *physiological* measures such as the galvanic skin response, the basis of the polygraph (lie detector) machine. More and more tests are now available for computer administration and this may become a popular category of tests.

Item structure. Another way to classify tests, which overlaps with the above, is through their item structure. Test items can be placed on a continuum from objective to subjective. At the objective end we have multiple-choice items; at the subjective end we have the type of open-ended questions that clinical psychologists and psychiatrists ask, such as "tell me more," "how do you feel about that?" and "tell me about yourself." In between we have countless variations such as matching items (closer to the objective pole) and essay questions (closer to the subjective pole). Objective items are easy to score and to manipulate statistically, but individually reveal little other than that the person answered correctly or incorrectly. Subjective items are difficult and sometimes impossible to quantify, but can be quite a revealing and rich source of information.

Another possible distinction in item structure is whether the items are *verbal* in nature or require *performance*. Vocabulary and math items are labelled verbal since they are composed of verbal elements; building a block tower is a performance item.

Area of Assessment. Tests can also be classified according to the area of assessment. For example, there are intelligence tests, personality questionnaires, tests of achievement, career-interest tests, tests of reading, tests of neuropsychological functioning, and so on. The MMY uses 16 such categories. These are not necessarily mutually exclusive categories, and many of them can be further subdivided. For example, tests of personality could be further categorized into introversion-extraversion, leadership, masculinity-femininity, and so on.

In this textbook we will look at these five major categories of tests:

1. Personality tests, which have played a major role in the development of psychological testing, both in its acceptance and criticism. Personality represents a major area of human functioning for social-behavioral scientists and lay persons alike;
2. Tests of cognitive abilities, not only traditional intelligence tests, but other dimensions of cognitive or intellectual functioning. In some ways, cognitive psychology represents a major new emphasis in psychology which has had a significant impact on all aspects of psychology both as a science and as an applied field;
3. Tests of attitudes, values, and interests, three areas that psychometrically overlap, and also offer lots of basic testing lessons;
4. Tests of psychopathology, primarily those used by clinicians and researchers to study the field of mental illness; and
5. Tests that assess normal and positive functioning, such as creativity, competence, and self-esteem.

Test function. Tests can also be categorized depending upon their function. Some tests are used to *diagnose* present conditions. (Does the client have a character disorder? Is the client depressed?) Other tests are used to make *predictions*. (Will this person do well in college? Is this client likely to attempt suicide?) Other tests are used in *selection* procedures, which basically involve accepting or not accepting a candidate, as in admission to graduate school. Some tests are used for *placement* purposes—candidates who have been accepted are placed in a particular "treatment." For example, entering students at a university may be placed in different level writing courses depending upon their performance on a writing exam. A battery of tests may be used to make such a placement decision or to assess which of several alternatives is most appropriate for the particular client—here the term typically used is *classification* (note that this term has both a broader meaning and a narrower meaning, see page 2). Some tests are used for *screening* purposes; the term screening implies a rapid and rough procedure. Some tests are used for *certification*, usually related to some legal standard; thus passing a driving test certifies that the person has at the very least a minimum proficiency, and is allowed to drive an automobile.

Score interpretation. Yet another classification can be developed on the basis of how scores on a test are interpreted. We can compare the score that an individual obtains with the scores of a group of individuals who also took the same test. This is called a *norm-reference*, since we refer to norms to give a particular score meaning; for most tests, scores are interpreted in this manner. We can also give meaning to a score by comparing that score to a decision rule called a *criterion*, so this would be a *criterion-reference*. For example, when you took a driving test (either written and/or road), the examiner did not say "Congratulations your score is two standard deviations above the mean." You either passed or failed based upon some predetermined criterion that may or may not have been explicitly stated. Note that norm-reference and criterion-reference refer not to the test but to how the score or performance is interpreted. The same test could yield either or both score interpretations.

Another distinction that can be made is whether the measurement provided by the test is *normative* or *ipsative*, that is, whether the standard of comparison reflects the behavior of others or of the client. Consider a 100-item vocabulary test which we administer to Marisa, and she obtains a score of 82. To make sense of that score we compare her score to some normative data—for example, the average score of similar-age college students. Now consider a questionnaire that asks Marisa to decide which of two values is more important to her: "Is it more important for

此为试读,需要完整PDF请访问: www.ertongbook.com