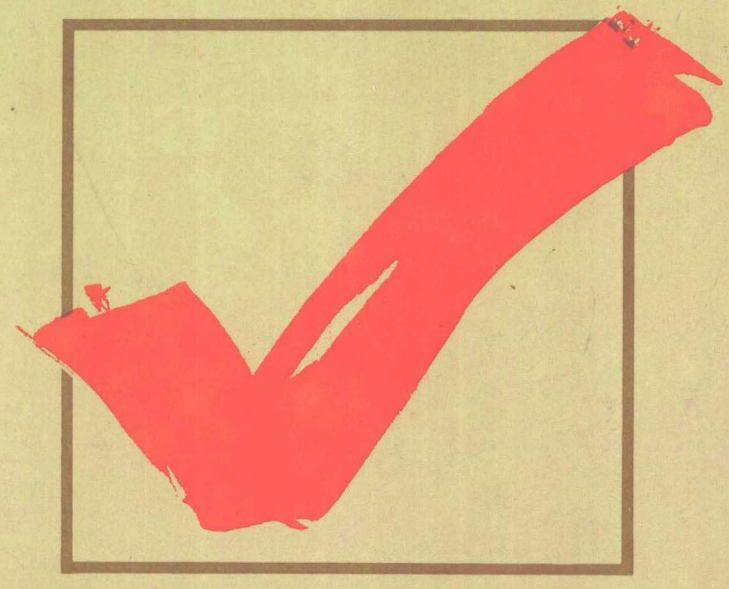
Handbook of Psychological Assessment



Gary Groth-Marnat

HANDBOOK OF PSYCHOLOGICAL ASSESSMENT

Gary Groth-Marnat


VAN NOSTRAND REINHOLD COMPANY

Library of Congress Catalog Card Number: 83-21812
ISBN: 0-442-22927-5

Manufactured in the United States of America

Published by Van Nostrand Reinhold Company Inc.
135 West 50th Street
New York, New York 10020

Van Nostrand Reinhold Company Limited
Molly Millars Lane
Wokingham, Berkshire RG11 2PY, England

Van Nostrand Reinhold
480 Latrobe Street
Melbourne, Victoria 3000, Australia

Macmillan of Canada
Division of Gage Publishing Limited
164 Commander Boulevard
Agincourt, Ontario M1S 3C7, Canada

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging in Publication Data

Groth-Marnat, Gary.
Handbook of psychological assessment.

Bibliography: p.
Includes index.
1. Psychological tests. I. Title. [DNLM:
1. Psychological tests – Handbooks. BF 176 G881h]
BF176.G76 1984 150'.28'7 83-21812
ISBN 0-442-22927-5

Preface

For many years I wondered why a composite book had never been written on interpreting the major psychological tests. This wondering began in graduate school with volumes of xeroxed manuals to accompany Anastasi and numerous expensive, often verbose texts on a single test. It became relatively easy to ignore the lack of such a book until I decided to teach a graduate course on assessment. There was simply no single book that I felt was adequate to address applied approaches to assessment. It was that dilemma which led to the writing of this book.

It is my hope that my colleagues and students will find the *Handbook of Psychological Assessment* to be a helpful, if not at times essential, aid in psychological assessments. In particular, I have tried to avoid the limitations of "cookbook" assessment procedures. Throughout the text, there has been continual reference to interrelations between scores, the importance of consulting outside sources, appropriate cautions relating to test data, defining the ideal role of the clinician, and the essential contribution that the client's personal history and observations of the client's behavior make towards the overall assessment procedure. It is hoped that this book will prove to be brief and succinct, but also that it will provide the tools necessary for developing a description of individuals that contains depth, accuracy, and clinical usefulness.

Since the book so often has seemed like a monster that just kept growing (and indeed is still hungry), I would like to make this preface brief and conclude with a tribute to those who were crucial, important, or helpful in my writing. The most important person is my original co-inspirer and co-conspirer, Dorothy Morena. Her initial help in editing and providing valuable ideas, rough drafts of the projective drawing chapter, and her overall integrity and warmth will always be remembered. Dayle Goldie also supplied much of the initial impetus, encouragement, and support. I would like to thank Mel Schwartz, Rick Thomas, Robert Zussman, George Sargent, and innumerable students for their reading of (and ideas regarding) different portions of the manuscript. Each of them, in his own way, served to correct my all-too-frequent nearsightedness and selective perception. I also greatly appreciate the two psychological evaluations contributed by George Sargent and Tom MacSpeiden. Finally, I would like to thank those persons who were directly involved in the preparation of the manuscript – especially Virginia Webster, who has aptly been described as "Wonder Woman" for her remarkable ability to mind read what to most people would appear to be unintelligible scribbles. More than just appreciation goes to Julie Fallscheer for her patience, tolerance of my mental absence, and preparation of the major portion of the reference section.

GARY GROTH-MARNAT

Contents

1
Introduction

The general purpose of the *Handbook of Psychological Assessment* is to provide a reference and instructional guide for professionals and students who are conducting psychological assessments. As a reference book, it can aid in the development of a large number and variety of interpretive hypotheses. It can also serve to point out obscure signs which are only infrequently encountered during evaluations but may still be crucial in developing a complete description of the client. As an instructional text, it provides the student with the basic tools relevant for interpreting the more frequently used psychological tests. Thus, it can relieve instructors of the need to cover a long discussion of interpretive hypotheses, thereby enabling them to devote more time to the evaluation and integration of test data. This book also provides a framework within which one can approach psychological test data in a coherent, problem-oriented manner. The goal is to aid the clinician and student in developing and integrating a wide variety of interpretive hypotheses within the context of a client's history, behavioral observations, and test data.

One significant and overriding emphasis in this book is its focus on assessing areas which are of practical use in evaluating individuals. It is applied in its orientation, and for the most part, theoretical discussions have been avoided. Many books written on psychological testing, as well as courses organized around these books, focus primarily on test theory and construction, and minimize actual interpretation of tests. In contrast, the intent of this book is to examine relevant features of test construction in such a way as to aid in evaluating psychological tests. Furthermore, the main approaches towards interpreting test data, both in a test battery and from individual tests, are outlined and elaborated upon.

However, the book was organized so as to provide a relatively brief discussion of approaches towards interpreting the main tests used in clinical practice. The danger of covering tests in such a relatively brief way is that clinicians, or especially students, may attempt to use the interpretive hypotheses in a "cookbook" fashion or treat the information from a "single sign" approach. For example, an examiner may read that large eyes on human figure drawings are a sign that the person is paranoid and may then not bother to integrate this finding with the other data. Thus, he may come to the possibly erroneous conclusion that

the single sign of drawing large eyes equals paranoia. It is clearly stressed throughout the book that clinicians should consider single test findings merely as hypotheses in need of further verification and that any finding should always be understood within the context of other sources of data. In order to facilitate more valid and useful interpretations, therefore, cautions and guidelines are provided for evaluating clinical and psychometric data. This book should not be used merely to interpret test scores as much as to assess individual persons within their unique situations. Test scores, then, are one tool to be utilized in the overall assessment process.

A further area of emphasis is that students/clinicians should be well aware of the assets and, perhaps more important, the limitations of each of the tests. Within a larger context, they should likewise be aware of the assets and limitations of testing as a general strategy in understanding people. Many tests are misused and students of testing should be clearly aware of the ways in which they are most likely to misuse testing. In a later section in this chapter, some of the precautions which must be taken prior to, and during, test use are discussed. Furthermore, each chapter on psychological tests contains a section dealing with the assets and limitations of each test. It is only as a result of a thorough understanding of the limitations of each test that its specific strengths can be maximized.

ORGANIZATION OF THE HANDBOOK

The decision for the inclusion of various tests was based primarily on their being the most frequently used tools for assessment. For example, according to several studies (Crenshaw, Bohn, Hoffman, Matheus, and Offenbach, 1968; Sundberg, 1961; Wade and Baker, 1977), the six most commonly used tests are: the Wechsler scales, the Rorschach, projective drawings (Draw-A-Person, House-Tree-Person), the Bender Visual Motor Gestalt (Bender), the Thematic Apperception Test (TAT), and the Stanford-Binet. The Wechsler Adult Intelligence Scale – Revised (WAIS-R) and the Wechsler Intelligence Scale for Children – Revised (WISC-R) were chosen for inclusion instead of the Stanford-Binet, because the WAIS-R and WISC-R not only are useful in providing cognitive assessments but also are helpful in evaluating personality and providing useful clinical information. In addition, both the Minnesota Multiphasic Personality Inventory (MMPI) and the California Personality Inventory (CPI) were chosen for inclusion. The MMPI is the most frequently used personality inventory, and of all the tests, it has the highest number of studies that have been conducted on it or performed using it (Alker, 1978). Finally, the CPI was selected because of its excellent technical development (Anastasi, 1982), numerous research studies, and relatively high frequency of use.

The different chapters in this book follow the general steps clinicians take when conducting evaluations. Chapter 1 deals with the preliminary issues that

clinicians must face prior to administering the tests. They must clarify their general role, evaluate the tests whose use they are considering, understand the setting in which they are working, and be aware of some of the major guidelines and cautions for using tests. It is also important for clinicians to have a general conceptual knowledge of the different phases of clinical assessment, beginning with a clarification of the problem and ending with the interpretation and integration of the data.

The middle part of the book (Chapters 2–8) provides a general working knowledge of the seven most frequently used tests. Each of these chapters begins with an introduction to the test in the form of a discussion of its history and development, current evaluation, and procedures for administration. The main portions of these chapters provide a guide towards interpretation which includes such areas as a discussion of the meaning of different scales, significant relations between scales, frequent trends which may be encountered, and the meaning of unusually high or low scores. When appropriate, additional subsections have been included. For example, the chapter on the Wechsler scales includes a discussion of the nature of intelligence because it is especially crucial for a clinician to understand the theoretical construct of "intelligence" prior to attempting an interpretation of I.Q. scores. Likewise, the chapter on the Thematic Apperception Test includes a summary of Murray's theory of personality because a knowledge of his concepts is also a necessary prerequisite for adequately understanding and interpreting the test.

The final step a clinician must take is the actual writing of the psychological report. Chapter 9 provides general guidelines for report writing, a report format, and four sample reports. The sample reports are representative of the four more common types of reports (psychiatric, legal, academic, personality) from the four most frequently encountered referral settings (medical setting, legal context, educational context, psychological clinic). It is hoped that the sequence will be a logical one for clinicians to follow and that the knowledge provided within the chapters will be useful, concise, and practical.

ROLE OF THE CLINICIAN

The central role of clinicians conducting assessments should be to answer specific questions and aid in making relevant decisions. This requires that clinicians be able to integrate a wide variety of data and bring into focus diverse areas of knowledge. Thus, they are not merely administering and scoring tests. A useful distinction which highlights this point is the contrast that Maloney and Ward (1976) have made between a clinician conducting psychological assessment and a psychometrist. Psychometrists tend more to use tests merely to obtain data, and their task is often perceived as emphasizing the clerical and technical aspects of testing. Their approach is primarily data oriented, and the end product is often

a series of traits or ability descriptions. These descriptions are typically not related to the person's overall context and do not address unique problems the person may be facing. In contrast, psychological assessment attempts to evaluate an individual in a problem situation so that the information derived from the assessment can somehow help with the problem. Tests are only one method of gathering data, and the test data are not end products, but merely tools, in the overall process of assessment. Psychological assessment, then, places data in a wide perspective; with its main focus being problem solving and decision making.

The distinction between psychometric testing and psychological assessment can be better understood, and the ideal role of the clinician more clearly defined, by briefly elaborating on the historical and methodological reasons for the development of the psychometric approach. When psychological tests were originally developed, there was an early and noteworthy success of group measurements of intelligence. This was especially true in military and industrial settings where individual interviewing and case histories were too expensive and time consuming for general use. The data-oriented intelligence tests were considered to be advantageous because they appeared "objective" and thus seemed to reduce possible interviewer bias. More important, they were quite successful in producing a relatively high number of true positives when used for classification purposes. Their predictions were generally accurate and usable. However, this created the early expectation that all assessments could be performed using the same method and would provide a similar level of accuracy and usefulness. Later assessment strategies often tried to imitate the methods of earlier intelligence tests for such variables as personality and psychiatric diagnosis.

A further development consistent with the psychometric approach was the concept of the "test battery." It was reasoned that if a single test could produce accurate descriptions of an ability or trait, then a series of tests could be administered to create a "total picture" of the person. The goal then was to develop a global, yet quantitative, description of the person using purely objective methods. This encouraged the idea that the tool (psychological test) was the best process for achieving this goal, rather than being merely one technique in the overall assessment procedure. Behind this approach were the concepts of individual differences and "trait" psychology. These assume that one of the best ways of describing the differences among individuals is to measure their strengths and weaknesses with respect to various traits. Thus, the clearest approach to the study of personality involved developing a relevant taxonomy of traits and then creating tests to measure these traits. Again, there was an emphasis on the tools as primary, with a de-emphasis on the input of the clinician. The result of these trends was a bias towards administration and clerical skills. Within this context, the psychometrist requires little, if any, clinical expertise other than in administering, scoring, and interpreting tests. Thus, the most preferred tests would be those that are machine scored, true-false, or multiple choice, and

which are constructed so that the normed scores – rather than the psychometrist – provide the interpretation.

The objective psychometric approach is most appropriately applicable to ability tests such as those measuring intelligence or mechanical skills. However, its usefulness decreases when attempting to assess personality traits such as dependence, authoritarianism, or anxiety. Personality variables are far more complex and therefore need to be validated within the context of history, behavioral observations, and interpersonal relationships. For example, a T score of 75 on the MMPI scale 9 (mania) takes on an entirely different meaning for a highly functioning physician than for an individual with a poor history of work and interpersonal relationships. When the purely objective psychometric approach is used for the evaluation of problems in living (neurosis, psychosis, etc.), its usefulness is questionable.

It is in the understanding and evaluation of personality and especially of problems in living that the approach which has been labeled *psychological assessment* is most useful. This is because these issues involve a particular problem situation having to do with a specific individual. The central role of the clinician performing psychological assessment is that of an expert in human behavior who must deal with complex processes and understand test scores within the context of a person's life. He must have knowledge concerning problem areas and, on the basis of this knowledge, be able to form a general idea regarding behaviors to observe and areas in which to collect relevant data. This involves an awareness and appreciation of multiple causation, interactional influences, and multiple relationships. As Woody (1980) has stated, "Clinical assessment is individually oriented, but it always considers social existence; the objective is usually to help the person solve problems."

In addition to an awareness of the role suggested by psychological assessment, there are also other specific areas of knowledge that clinicians should have. These include personality theory, abnormal psychology, and the psychology of adjustment, as well as a knowledge of test construction and of basic statistics. Furthermore, they should know the main interpretive hypotheses in psychological testing and be able to identify, sift through, and evaluate a series of hypotheses and decide on the most relevant ones. For each assessment device, it is also important that clinicians have a conception of what it is they are trying to test. Thus, rather than merely knowing the labels and definitions for various types of anxiety or thought disorders, they should also have in-depth operational criteria for them. For example, the concept of intelligence, as represented by the I.Q. score, can sometimes appear misleadingly straightforward. However, intelligence test scores can be complex, involving a variety of different cognitive abilities, the influence of cultural factors, varying performance under different conditions, and issues related to the nature of intelligence. Unless clinicians are familiar with these areas, they are not adequately prepared to handle I.Q. data.

A problem encountered in many training programs is that although students will frequently have a knowledge of abnormal psychology, personality theory, and test construction, they are usually not sufficiently trained in how to integrate these knowledge areas into the interpretation of test results. Instead, their training focuses on developing competency in administration and scoring, rather than on knowledge relating to what it is that they are testing.

The role stressed in this book is consistent with the approach of psychological assessment in that the clinician not only should be knowledgeable in regard to traditional content areas in psychology and the nature of what is being tested, but also should be able to integrate the test data into a relevant description of the person. This description, although focusing on the individual, should be able to take into account the complexity of his social environment, personal history, and behavioral observations. Yet, the end goal is not merely to describe the person, but rather to develop relevant answers to specific questions, aid in problem solving, and facilitate decision making.

EVALUATING PSYCHOLOGICAL TESTS

Prior to using a psychological test, a clinician should investigate and understand the theoretical orientation of the test, practical considerations, the appropriateness of the standardization sample, and the adequacy of its reliability and validity. Often, extremely helpful reviews which relate to these issues can be found in current and past editions of Buros's *Mental Measurement Yearbook.* Table 1-1 outlines the more important questions which have to be answered. The issues outlined in this table will be discussed further. The discussion is consistent with the practical orientation of this text in that it focuses on problems which clinicians using psychological tests are likely to confront. It is not intended to provide a comprehensive coverage of test theory and construction; if a more detailed treatment is required, the reader is referred to one of the many texts on psychological testing (e.g., Anastasi, 1982; Kaplan and Sacuzzo, 1982).

Theoretical Orientation

One of the foremost requirements prior to evaluating whether or not a test is appropriate is for examiners to understand the theoretical orientation of the test. They should research the construct that the test is supposed to measure and then examine the manner in which the test approaches this construct. This information can usually be found in the test manual. However, if for any reason the information in the manual is not sufficient, then further knowledge should be sought elsewhere. Clinicians can frequently obtain useful information regarding the trait by assessing the individual test items. Usually an individual analysis of the items, which can help the potential test user to evaluate whether or not these items appear relevant to the trait being measured, can be found in the manual.

Table 1-1. Issues to Address When Evaluating a Psychological Test.

Theoretical Orientation

1. Do you adequately understand the theoretical construct the test is supposed to be measuring?
2. Do the test items correspond to the theoretical description of the construct?

Practical Considerations

1. If reading is required by the examinee, does his ability match the level required by the test?
2. How appropriate is the length of the test?
3. Does the examiner require additional training? If so, how can this be acquired?

Standardization

1. Is the population to be tested similar to the population the test was standardized on?
2. Was the size of the standardization sample adequate?
3. Have specialized subgroup norms been developed?

Reliability

1. Are reliability estimates sufficiently high (generally .90 for clinical decision making and .70 for research purposes)?
2. What implications do the relative stability of the trait, the method of estimating reliability, and the test format have on reliability?

Validity

1. What were the criteria and procedures used to validate the test?
2. Has the test been constructed so as to produce accurate measurements?
3. Will the test produce accurate measurements within the context and purpose for which you would like to use it?

Practical Considerations

There are a number of practical issues which do not relate as much to the construction of the test as to the context and manner in which the test will be used. First of all, tests vary in terms of the degree of education which examinees must have in order to understand them adequately. This may be especially important in relation to any reading which is demanded of the examinee. The examinee must be able to read, comprehend, and respond appropriately to the test. Some tests are too long, which can lead to a loss of rapport with (or extensive frustration on the part of) the examinee. Sometimes this problem can be reduced by administering short forms of the test, provided these short forms have been properly developed and are treated with appropriate caution. Finally, clinicians have to assess the extent to which they will need training to administer and interpret the instrument. If further training is necessary, then a plan must be developed for acquiring this training.

Standardization

Yet, another central issue relates to the adequacy of norms. Each test has norms that reflect the distribution of scores by a specific standardization sample. The basis upon which individual test scores have meaning relates directly to the similarity between that individual and the standardization sample. If there is a similarity between the group or individual being tested and the standardization sample, then adequate comparisons can be made. For example, if the test was standardized on college students between the ages of 18 and 22, then, if one assumes that the test is otherwise sufficiently reliable and valid, useful comparisons can be made for college students within that age bracket. The more dissimilar the person is from this standardization group (e.g., over 70 years of age with low educational achievement), the less useful the test is in evaluating him. The examiner may need to consult the literature to see if research which followed the publication of the test manual has developed norms for different groups. This is particularly important for tests such as the MMPI and the Rorschach where norms for children and adolescents have recently been published.

There are three major questions relating to the adequacy of norms which must be answered. The first is whether or not the standardization group is representative of the population on which the examiner would like to use the test. The test manual should include sufficient information to determine the representativeness of the standardization sample. If this information is not sufficient or is in any way incomplete, then the degree of confidence with which the test can be used is greatly reduced. The ideal and current practice is to use stratified random sampling. However, this can be an extremely costly and time consuming procedure, and as a result, many tests are grossly deficient in this respect. The second question is whether the size of the standardization group is large enough. If the group is too small, then the results may not give stable estimates because there may be too much random fluctuation within the group. Finally, a good test will have specialized subgroup norms as well as broad national norms. Knowledge relating to subgroup norms will give examiners greater flexibility and confidence if they are using the test with similar subgroup populations. This is particularly important when subgroups produce significantly different sets of scores from the normal standardization group. These subgroups can be based on such factors as sex, geographic location, age, level of education, socioeconomic status, or urban versus rural environment. Knowledge of each of these subgroup norms allows for more appropriate and meaningful interpretations of scores.

Reliability

The reliability of a test refers to its degree of stability, consistency, and accuracy. In other words, it addresses the question of the extent to which scores obtained

by a person will be the same if the person is reexamined by the same test on different occasions. Underlying the concept of reliability is the possible range of error or "error of measurement" of a single score. This is an estimate of the range of possible random fluctuation which can be expected in an individual's score. However, it should be stressed that there will always be a certain degree of error or "noise" in the system resulting from such factors as misreading of the items, poor administration procedures, or the changing mood of the client. If there is a large degree of random fluctuation, then the examiner cannot place a great deal of confidence in an individual's scores. It is the goal of the test constructor to reduce the degree of measurement error, or random fluctuation, as much as possible. If this is achieved, then the difference between one score and another is more likely to be due to some true difference in the characteristic being measured rather than some chance fluctuation.

There are two main issues relating to the degree of error in a test. The first is that there is an inevitable, natural variation in human performance. Usually the variability is less for measurements of ability than for those of personality. Whereas ability variables (intelligence, mechanical aptitude, etc.) show gradual changes resulting from growth and development, many personality traits are much more highly dependent on factors such as mood. This is particularly true in the case of a characteristic such as anxiety. The practical significance of this in evaluating a test is that certain factors outside the test itself can serve to reduce the reliability which the test can realistically be expected to achieve. Thus, an examiner should generally expect higher reliabilities for an intelligence test than for a test measuring a personality variable such as anxiety. This makes it the examiner's responsibility to know the nature of that which is being measured, especially with regard to the degree of variability that is to be expected in the trait being measured.

The second important issue relating to reliability is that psychological testing methods are necessarily imprecise. Within the fields of the "hard" sciences, a direct measurement can be made such as the concentration of a chemical solution, the relative weight of one organism compared to another, or the strength of radiation. In contrast to this are the constructs in psychology where measurements are often indirect. We cannot perceive something such as "intelligence" directly; rather, we infer its existence by measuring behavior we have defined as being intelligent. Whereas it is not possible to control for the natural variability in human performance, adequate test construction can attempt to reduce the degree of imprecision which is a function of the test. Both natural human variability and test imprecision make the task of measurement extremely difficult. Although some error in testing is inevitable, the goal of test construction is to keep it within reasonably accepted limits. A "high" correlation is generally .80 or more, but the variable being measured will also change the expected strength of the correlation. Likewise, the method of determining reliability will alter the

relative strength of the correlation. Usually clinicians should look for correlations of .90 or higher in tests that will be used to make decisions about individuals, whereas a correlation of .70 or more is generally adequate for research purposes.

The purpose of reliability is to estimate the degree to which the test varies due to error. There are three primary methods of obtaining reliability. Reliability can refer to the extent to which the test produces consistent results upon retesting (test-retest), the relative accuracy of a test at a given time (alternate forms), and the internal consistency of the items (split half). Another way of summarizing this is that reliability can be time to time (test-retest), form to form (alternate forms), or item to item (split half). Whereas these are the main types of reliability, there is a fourth one – the Kuder-Richardson – which, like the split half, is a measurement of the internal consistency of the test items. However, this method is considered appropriate only for tests which are relatively pure measures of a single variable, and it will not be covered here.

Test-Retest Reliability. Test-retest reliability is determined by administering the test and then giving a repeat administration on a second occasion. The reliability coefficient is determined by correlating the scores obtained by the same person on the two different administrations. The degree of correlation between the two scores indicates to what extent the test can be generalized from one situation to the next. If the correlations are high, then the results are less likely to be due to random fluctuations in the condition of the examinee or the testing environment. Thus, the examiner can be relatively confident that differences in scores are the result of an actual change in the trait being measured.

There are a number of factors which must be considered in assessing the appropriateness of test-retest reliability. One is that the interval between administrations can affect reliability. Thus, a test manual should clearly specify the interval as well as any significant life changes that the examinees may have experienced such as counseling, career changes, or psychotherapy. For example, tests of preschool intelligence often give reasonably high correlations if the second administration is within several months of the first one. However, correlations with later childhood or adult I.Q.'s are generally useless because of the many intervening life experiences. One of the major difficulties with test-retest reliability is the effect that practice and memory may have on performance which can produce improvement between one administration and the next. This is a particular problem for speeded and memory tests such as those found on the Digit Symbol and Arithmetic subtests of the WAIS. Additional sources of error may be the result of random, short-term fluctuations in the examinee or variations in the testing conditions. In general, test-retest reliability is the preferred method only if the variable being measured is relatively stable. If the variable is a highly changeable one such as anxiety, then this method is usually not adequate.