

## CHAPTER 1

# The Assessment Process

IRVING B. WEINER

<b>COLLECTING ASSESSMENT INFORMATION</b>	<b>3</b>
<b>INTERPRETING ASSESSMENT INFORMATION</b>	<b>11</b>

<b>USING ASSESSMENT INFORMATION</b>	<b>20</b>
<b>REFERENCES</b>	<b>23</b>

Assessment psychology is the field of behavioral science concerned with methods of identifying similarities and differences among people in their psychological characteristics and capacities. As such, psychological assessment comprises a variety of procedures that are employed in diverse ways to achieve numerous purposes. Assessment has sometimes been equated with testing, but there is more to the assessment process than giving tests. Psychological assessment involves integrating information gleaned not only from test protocols but also from interviews, behavioral observations, collateral reports, and historical documents. In this regard, the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 1999) specify that

the use of tests provides one method of collecting information within the larger framework of a psychological assessment of an individual. . . . A psychological assessment is a comprehensive examination undertaken to answer specific questions about a client's psychological functioning during a particular time interval or to predict a client's psychological functioning in the future. (p. 119)

The diverse ways in which assessment procedures are employed include many alternative approaches to obtaining and combining information from different sources, and the numerous purposes that assessment serves arise in response to a broad range of referral questions raised in clinical, educational, health care, forensic, and organizational settings. Chapters in this volume elaborate the diversity of assessment procedures, the nature of the

assessment questions that arise in various settings, and the types of assessment methods commonly employed to address these questions.

This introductory chapter sets the stage for what is to follow by conceptualizing assessment as a three-stage process of information input, information evaluation, and information output. Information input involves collecting assessment data of appropriate kinds and in sufficient amounts to address referral questions in meaningful and useful ways. Information evaluation consists of interpreting assessment data in a manner that provides accurate descriptions of respondents' psychological characteristics and behavioral tendencies. Information output calls for using descriptions of respondents to formulate conclusions and recommendations that help to answer referral questions. Each of these phases of the assessment process requires assessors to accomplish some distinctive tasks, and each involves choices and decisions that touch on critical issues in conducting psychological assessments.

### COLLECTING ASSESSMENT INFORMATION

The process of collecting assessment information begins with a formulation of the purposes that the assessment is intended to serve. Knowing why an assessment is being conducted helps examiners identify tests and other sources of information that are likely to provide a basis for relevant conclusions and useful recommendations. Also helpful in planning the data collection process is attention to examiner, respondent, and data management characteristics that influence the nature and utility of whatever findings are obtained.

## 4 Assessment Issues

### Formulating Goals

Psychological assessments are instigated by referrals that pose questions about aspects of a person's psychological functioning or likely future behavior. Clearly stated and psychologically relevant referral questions help psychologists determine what kinds of assessment data to collect, what considerations to address in examining these data, and what implications of their findings to emphasize in their reports. Referral questions that lack clarity or psychological relevance require some reformulation to give direction to the assessment process. For example, a referral in a clinical setting that asks vaguely for personality evaluation or differential diagnosis needs to be made more specific in consultation between the psychologist and the referring person to identify why a personality evaluation is being sought or what diagnostic possibilities are at issue. Assessment in the absence of a specific referral question can be a sterile exercise in which neither the data collection process nor the psychologist's inferences can be focused in a meaningful way.

Even when referral questions are adequately specified, they may not be psychological in nature. Assessors doing forensic work are often asked to evaluate whether a criminal defendant was insane at the time of an alleged offense; however, *sanity* is a legal term, not a psychological term. No assessment methods are designed to identify "insanity," nor are there any research studies in which "being insane" has been used as an independent variable. To help assessors plan their procedures and frame their reports in such instances, the referral must be translated into psychological terms, as in defining *insanity* as the inability to distinguish reality from fantasy.

As a further challenge in formulating assessment goals, referral questions that are specific and framed in psychological terms may still be unclear if they concern complex and multidetermined patterns of behavior. In employment evaluations, for example, a referring person may want to know which of three individuals is likely to perform best in a position of leadership or executive responsibility. To address this type of question adequately, assessors must first be able to identify psychological characteristics that are likely to make a difference in the particular circumstances, as by proceeding, in this example, with the presumption that being energetic, decisive, assertive, self-confident, and relatively unflappable is likely to contribute to effective leadership. The data collection process can then be planned to measure these characteristics, and the psychologist's eventual report can specify their presence or absence as the basis for whatever hiring decision is recommended.

### Selecting Tests

The multiple sources of assessment information mentioned include:

- the results of formal psychological testing with standardized instruments;
- responses to questions asked in structured and unstructured interviews;
- observations of behavior in various types of contrived situations and natural settings;
- reports from relatives, friends, employers, and other collateral persons concerning an individual's previous life history and current characteristics and behavioral tendencies; and
- documents such as medical records, school records, and written reports of earlier assessments.

Individual assessments vary considerably in the availability and utility of these diverse sources of information. Assessments may at times be based solely on record reviews and collateral reports, because the person being assessed is unwilling to be seen by an examiner or is for some reason prevented from doing so. Some people being assessed are quite forthcoming when interviewed but are reluctant to be tested; others find it difficult to talk about themselves but are quite responsive to testing procedures; and in still other cases, in which both interview and test data are ample, there may be a dearth of collateral input or documented information on which to draw.

There is little way to know before the fact which sources of information will prove most critical or valuable in an assessment process. What collateral informants say about a person may sometimes be more revealing and reliable than what the person says about him- or herself, and in some instances historical documents may prove more informative and dependable than either first-person or collateral reports. Behavioral observations and interview data may on occasion contribute more to an assessment than standardized tests or may even render testing superfluous, whereas in other instances formal psychological testing may reveal vital diagnostic information that would otherwise not have been uncovered.

The fact that psychological assessment can proceed effectively without psychological testing helps to distinguish between these two activities. As previously noted, the terms *psychological assessment* and *psychological testing* are sometimes used synonymously, but psychological testing is only one of several sources of information typically utilized in the assessment process. (See Matarazzo, 1990; Meyer et al., 2001.) Testing, however,

is the most complex and specialized of the data collection procedures used in psychological assessment, starting with the selection of an appropriate test battery from among an extensive array of available measuring instruments. (See Geisinger, Spies, Carlson, & Plake, 2007; see also Ben-Porath; Lowman and Carson; O'Brien and Young; Otero, Podell, DeFina and Goldberg; Viglione and Rivera; and Wasserman, this volume). The chief considerations that should determine the composition of a test battery are:

- the psychometric adequacy of the measures being considered;
- the relevance of these measures to the referral questions being addressed;
- the likelihood that these measures will contribute incremental validity to the decision-making process; and
- the additive, confirmatory, and complementary functions that individual measures are likely to serve when used jointly.

#### ***Psychometric Adequacy***

As elaborated in the *Standards for Educational and Psychological Testing* (AERA et al., 1999) and in Wasserman and Bracken, this volume, the psychometric adequacy of an assessment instrument consists of the extent to which it

- involves standardized test materials and administration procedures;
- can be coded with reasonably good interscorer agreement;
- demonstrates acceptable reliability;
- has generated relevant normative data; and
- shows valid corollaries that serve the purposes for which it is intended.

Assessment psychologists may at times choose to use tests with uncertain psychometric properties, perhaps for exploratory purposes or for comparison with a previous examination in which these tests have been used. Generally speaking, however, formal testing as part of a psychological assessment should be limited to standardized, reliable, and valid instruments for which there are adequate normative data.

#### ***Relevance***

The tests selected for inclusion in an assessment battery should provide information relevant to answering the questions that have been raised about the person being examined. Questions that relate to personality functions (e.g., “What kind of approach in psychotherapy is likely

to be helpful to this person?”) call for personality tests. Questions that relate to educational issues (e.g., “Does this student have a learning disability?”) call for measures of intellectual abilities and academic aptitude and achievement. Questions that relate to neuropsychological functions (e.g., “Are there indications of memory loss?”) call for measures of cognitive functioning, with special emphasis on measures of capacities for learning and recall.

These examples of relevance may seem too obvious to have warranted mention. However, they reflect an important and sometimes overlooked guiding principle in psychological assessment, namely, that test selection should be justifiable for each measure included in an assessment battery. Insufficient attention to justifying the use of particular measures in specific instances can result in two ill-advised assessment practices: (1) conducting examinations with a fixed and unvarying battery of measures regardless of what questions are being asked in the individual case, and (2) using favorite instruments at every opportunity even when they are unlikely to serve any central or unique purpose in a particular assessment.

The administration of minimally useful tests that have little relevance to the referral question is a wasteful procedure that can evoke deserved criticism of assessment psychologists and the assessment process. Likewise, the propriety of charging fees for unnecessary procedures can rightfully be challenged by persons receiving or paying for these services, and the competence of assessors who give tests that make little contribution to answering the questions at issue can be challenged in such public forums as the courtroom. (See Weiner, 2009.)

#### ***Incremental Validity***

*Incremental validity* in psychological assessment refers to the extent to which additional information increases the accuracy of a classification or prediction above and beyond the accuracy achieved by information already available. Assessors pay adequate attention to incremental validity by collecting the amount and kinds of information they need to answer a referral question, but no more than that. In theory, then, familiarity with the incremental validity of various measures when used for certain purposes, combined with test selection based on this information, minimizes redundancy in psychological assessment and satisfies both professional and scientific requirements for justifiable test selection.

In practice, however, strict adherence to incremental validity guidelines often proves difficult and even disadvantageous to implement. As already noted, it is difficult to anticipate which sources of information will prove to

be most useful in an assessment. Similarly, with respect to which instruments to include in a test battery, there is little way to know whether the tests administered have yielded enough data and which tests have contributed most to understanding the person being examined, until after the data have been collected and analyzed.

In most practice settings, it is reasonable to conduct an interview and review previous records as a basis for deciding whether formal testing would be likely to help answer a referral question—that is, whether it will show enough incremental validity to warrant its cost in time and money. Likewise, reviewing a set of test data can provide a basis for determining what kind of additional testing might be worthwhile. However, it is rarely appropriate to administer only one test at a time, to choose each subsequent test on the basis of the preceding one, and to schedule a further testing session for each additional test administration. For this reason, responsible psychological assessment usually consists of one or two testing sessions comprising a battery of tests selected to serve specific additive, confirmatory, and complementary functions.

#### *Additive, Confirmatory, and Complementary Functions of Tests*

Some referral questions require selection of multiple tests to identify relatively distinct and independent aspects of a person's psychological functioning. For example, students receiving low grades may be referred for an evaluation to help determine whether their poor academic performance is due primarily to limited intelligence or to personality characteristics that are fostering negative attitudes toward achieving in school. A proper test battery in such a case would include some measure of intelligence and some measure of personality functioning. These two measures would then be used in an additive fashion to provide separate pieces of information, both of which would contribute to answering the referral question. As this example illustrates, the additive use of tests serves generally to broaden understanding of the person being examined.

Other assessment situations may create a need for confirmatory evidence supporting conclusions based on test findings, in which case two or more measures of the same psychological function may have a place in the test battery. Assessors conducting a neuropsychological examination to address possible onset of Alzheimer's disease, for example, ordinarily administer several memory tests. Should each of these tests identify memory impairment consistent with Alzheimer's, from a technical standpoint, only one of them would have been necessary, and the others have shown no incremental validity. Practically speaking,

however, the multiple memory measures taken together provide confirmatory evidence of memory loss. Such confirmatory use of tests strengthens understanding and helps assessors present conclusions with confidence. The confirmatory function of a multitest battery is especially useful when tests of the same psychological function measure it in different ways. The advantages of multimethod assessment of variables have long been recognized in psychology, beginning with the work of Campbell and Fiske (1959) and continuing with the reports of the American Psychological Association's (APA's) Psychological Assessment Work Group, which stress the improved validity that results when phenomena are measured from a variety of perspectives (Kubiszyn et al., 2000; Meyer et al., 2001):

The optimal methodology to enhance the construct validity of nomothetic research consists of combining data from multiple methods and multiple operational definitions. . . . Just as effective nomothetic research recognizes how validity is maximized when variables are measured by multiple methods, particularly when the methods produce meaningful discrepancies . . . the quality of idiographic assessment can be enhanced by clinicians who integrate the data from multiple methods of assessment. (Meyer et al., 2001, p. 150)

Such confirmatory testing is exemplified in applications of the Minnesota Multiphasic Personality Inventory (MMPI, MMPI-2) and the Rorschach Inkblot Method (RIM), the two most widely researched and frequently used personality assessment instruments (Butcher & Rouse, 1996; Hogan, 2005). As discussed later in this chapter and in Viglione and Rivera and in Ben-Porath, this volume, the MMPI-2 is a relatively structured self-report inventory, whereas the RIM is a relatively unstructured measure of perceptual-cognitive and associational processes. (See also Exner, 2003; Graham, 2006; Greene, 2011; Weiner, 2003.) Because of differences in their format, the MMPI-2 and the RIM measure normal and abnormal characteristics in different ways and at different levels of a person's ability and willingness to recognize and report them directly. Should a person display some type of disordered functioning on both the MMPI-2 and the RIM, this confirmatory finding becomes more powerful and convincing than having such information from one of these instruments but not the other, even though technically in this instance no incremental validity derives from the second instrument. (See Weiner, 2005.)

Confirmatory evidence of this kind often proves helpful in professional practice, especially in forensic work. As described by Heilbrun (2001; see also Heilbrun, Grisso, & Goldstein, 2009), multiple sources of information pointing

in the same direction bolster courtroom testimony, whereas conclusions based on only one measure of some characteristic can result in assessors being criticized for failing to conduct a thorough examination.

Should multiple measures of the same psychological characteristics yield different instead of confirmatory results, these results usually can serve valuable complementary functions in the interpretive process. At times, apparent lack of agreement between two purported measures of the same characteristic has been taken to indicate that one of the measures lacks convergent validity. This negative view of divergent test findings fails to take adequate account of the complexity of the information provided by multimethod assessment, and it can result in misleading conclusions. To continue with the example of conjoint MMPI-2 and RIM testing, suppose that a person's responses show elevation on indices of depression on one of these measures but not the other. Inasmuch as indices on both measures have demonstrated some validity in detecting features of depression, the key question to ask is not which measure is wrong in this instance but rather why the measures have diverged.

Perhaps, as one possible explanation, the respondent has some underlying depressive concerns that he or she does not recognize or prefers not to admit to others, in which case depressive features might be less likely to emerge in response to the self-report MMPI-2 methodology than on the more indirect Rorschach task. Or perhaps the respondent is not particularly depressed but wants very much to give the impression of being in distress and needing help, in which case the MMPI-2 might be more likely to show depression than the RIM. Or perhaps the person generally feels more relaxed and inclined to be forthcoming in relatively structured than relatively unstructured situations; then the MMPI-2 is more likely than the RIM to reveal whether the person is depressed.

As these examples show, multiple measures of the same psychological characteristic can complement each other when they diverge, with one measure sometimes identifying the presence of a characteristic (a true positive) that is missed by the other (a false negative). Possible reasons for the false negative can contribute valuable information about the respondent's test-taking attitudes and likelihood of behaving differently in situations that differ in the amount of structure they provide. The translation of such divergence between MMPI-2 and RIM findings into clinically useful diagnostic inferences and individual treatment planning has been elaborated by Finn (1996) and Ganellen (1996). Whatever measures may be involved, the complementary use of divergent test

findings frequently enriches the interpretive yield of an assessment process.

### Examiner Characteristics

Two characteristics of the examiners who conduct psychological assessments are likely to influence the amount and kind of data they collect. One of these characteristics consists of their qualifications and competence in using assessment procedures, and the other concerns ways in which their personal qualities may affect how different kinds of people respond to them.

#### *Qualifications and Competence*

There is general consensus that persons who conduct psychological assessments should be qualified by education and training to do so. The Ethical Principles and Code of Conduct promulgated by the APA (2002) offers the following guideline in this regard: "Psychologists provide services, teach, and conduct research with populations and in areas only within the boundaries of their competence, based on their education, training, supervised experience, consultation, study, or professional experience" (Ethical Code 2.04[a]). Particular kinds of knowledge and skill that are necessary for test users to conduct adequate assessments are specified further by Krishnamurthy and Yalof (2010) and in the test user qualifications endorsed by the APA (Turner, DeMers, Fox, & Reed, 2008). Of further note is the identification in the *Standards for Educational and Psychological Testing* (AERA et al., 1999) of who is responsible for the proper use of tests in psychological assessments: "The ultimate responsibility for appropriate test use and interpretation lies predominantly with the test user. In assuming this responsibility, the user must become knowledgeable about a test's appropriate uses and the populations for which it is suitable" (p. 112).

Despite the clarity of these statements and the detailed guidelines they contain, two persistent issues in contemporary assessment practice remain unresolved. The first issue is that, typically, adequate qualifications are assumed for psychologists who are licensed and present themselves as competent to provide psychological testing services. However, until such time as the criteria proposed in the test user qualifications become incorporated into licensing procedures, being licensed to use psychological tests does not ensure being competent in using them.

Being competent in psychological testing requires familiarity with the latest revision of whatever instruments an assessor is using, with current research and the most recent normative data concerning these instruments,

and with the manifold interpretive complexities they are likely to involve. Assessment competence also requires appreciation for a variety of psychometric, interpersonal, sociocultural, and contextual issues that affect not only the collection of assessment information but also its interpretation and utilization. The chapters that follow in this volume bear witness to the broad range of these issues and to the steady output of research findings, practice guidelines, and new or revised measures that make assessment psychology a dynamic and rapidly evolving field with a large and burgeoning literature. Only by keeping reasonably current with these developments can psychological assessors remain competent, and only by remaining competent can they meet their ethical responsibilities (Kitchener, 2000; Koocher & Keith-Spiegel, 2008; Weiner, 1989).

The second persistent issue concerns assessment by persons who are not psychologists and are therefore not bound by this profession's ethical principles or guidelines for practice. Nonpsychologist assessors who can obtain psychological tests are free to use them however they wish. When easily administered measures yield test scores that seem transparently interpretable, as in the case of an elevated Borderline scale on the Millon Multiaxial Clinical Inventory–III (MCMI–III; Millon, 2006) or an elevated Paranoia scale on the Personality Assessment Inventory (PAI; Morey, 2007), unqualified examiners can draw superficial conclusions that fail to consider the complexity of these instruments, the interactions among their scales, and the limits of their applicability. It accordingly behooves assessment psychologists not only to maintain their own competence but also to call attention, when necessary, to assessment practices that fall short of reasonable standards of competence.

### *Personal Influence*

Assessors can influence the information they collect by virtue of their personal qualities and by the manner in which they conduct a psychological examination. In the case of self-administered measures such as interest surveys or personality questionnaires, examiner influence may be minimal. Interviews and interactive testing procedures, however, create ample opportunity for an examiner's age, gender, ethnicity, or other characteristics to make respondents feel more or less comfortable and more or less inclined to be forthcoming. Examiners accordingly need to be alert to instances in which such personal qualities may be influencing the nature and amount of the data they are collecting.

The most important personal influence that examiners cannot modify or conceal is their language facility.

Psychological assessment procedures are extensively language-based, both in their content and in the instructions that introduce nonverbal tasks, and accurate communication is essential for obtaining reliable assessment information. It is widely agreed that both examiners and whomever they are interviewing or testing should be communicating either in their native language or in a second language in which they are highly proficient (AERA et al., 1999). The use of interpreters to circumvent language barriers in the assessment process rarely provides a satisfactory solution to this problem. Unless an interpreter is fully conversant with idiomatic expressions and cultural referents in both languages, is familiar with standard procedures in psychological assessment, and is a stranger to the examinee (as opposed to a friend, relative, or member of the same closely knit subcultural community), the obtained results may be of questionable validity.

Likewise, in the case of self-administered measures, instructions and test items must be written in a language that the respondent can be expected to understand fully. Translations of pencil-and-paper measures require close attention to the idiomatic vagaries of each new language and to culture-specific contents of individual test items, in order to ensure equivalence of measures in the cross-cultural applications of tests (Allen & Walsh, 2000).

Unlike their fixed qualities, examiners' manner in conducting the assessment process is within their control, and untoward examiner influence can be minimized by appropriate efforts to promote a full and open response to the assessment procedures. To achieve this end, an assessment typically begins with a review of its purposes, a description of the procedures that will be followed, and efforts to establish a rapport that will help the person being evaluated feel comfortable and willing to cooperate with the assessment process. Variations in examiner behavior while introducing and conducting psychological evaluations can substantially influence how respondents perceive the assessment situation—for example, whether they see it as an authoritarian investigative process intended to ferret out defects and weaknesses or as a mutually respectful and supportive interaction intended to provide understanding and help. Even while following closely the guidelines for a structured interview and adhering faithfully to standardized procedures for administering various tests, examiners need to recognize that their manner, tone of voice, and apparent attitude are likely to affect the perceptions and comfort level of the person being assessed and, consequently, the amount and kind of information that person provides. (See Masling, 1966, 1998.)

### Respondent Issues

Examiner influence in the assessment process inevitably interacts with the attitudes and inclinations of the person being examined. Some respondents may feel more comfortable being examined by an older person than a younger one, for example, or by a male than a female examiner, whereas other respondents may prefer a younger and female examiner. Among members of a minority group, some may prefer to be examined by a person with a cultural or ethnic background similar to theirs, whereas others are less concerned with the examiner's background than with his or her competence. Similarly with respect to examiner style, a passive, timid, and dependent person might feel comforted by a warm, friendly, and supportive examiner approach that would make an aloof, distant, and mistrustful person feel uneasy; conversely, an interpersonally cautious and detached respondent might feel safe and secure when being examined in an impersonal and businesslike manner that would be unsettling and anxiety provoking to an interpersonally needy and dependent respondent. With such possibilities in mind, skilled examiners usually vary their behavioral style with an eye to conducting assessments in ways that will be likely to maximize each individual respondent's level of comfort and cooperation.

Two other respondent issues that influence the data collection process concern a person's right to give informed consent to being evaluated and his or her specific attitudes toward being examined. With respect to informed consent, the introductory phase of conducting an assessment ordinarily must include not only the explanation of purposes and procedures mentioned previously, which informs the respondent, but also an explicit agreement by the respondent (or persons legally responsible for the respondent) to undergo the evaluation. As elaborated in the Standards for Educational and Psychological Testing (AERA et al., 1999), informed consent can be waived only when an assessment has been mandated by law, as in the case of a court-ordered evaluation, or when it is implicit, as when a person applies for a position for which being assessed is a requirement (i.e., a job for which all applicants are being screened psychologically; see also Kitchener, 2000, and the chapters by Geisinger & McCormick and by Rey-Casserly & Koocher, this volume).

Having given their consent to be evaluated, moreover, respondents are entitled to revoke it at any time during the assessment process. Hence, the prospects for obtaining adequate assessment data depend not only on whether respondents can be helped to feel comfortable and be forthcoming but even more basically on whether they

consent in the first place to being evaluated and remain willing during the course of the evaluation.

Issues involving a respondent's specific attitudes toward being examined typically arise in relation to whether the assessment is being conducted for clinical or for administrative purposes. When assessments are being conducted for clinical purposes, the examiner is responsible to the person being examined, the person being examined is seeking some type of assistance, and the examination is intended to be helpful to this person and responsive to his or her needs. As common examples in clinical assessments, people concerned about their psychological well-being may seek an evaluation to learn whether they need professional mental health care, and people uncertain about their educational or vocational plans may look for help in determining what their abilities and interests suit them to do.

In administrative assessments, by contrast, examiners are responsible not to the person being examined but to some third party who has requested the evaluation to assist in arriving at some judgment about the person. Examiners in an administrative assessment are ethically responsible for treating the respondent fairly and with respect, but the evaluation is being conducted for the benefit of the party requesting it, and the results may or may not meet the respondent's needs or serve his or her best interests. Assessment for administrative purposes occurs commonly in forensic, educational, and organizational settings when evaluations are requested to help decide such matters as whether a prison inmate should be paroled, a student should be admitted to a special program, or a job applicant should be hired. (See Heilbrun, 2001; Monahan, 1980.)

As for their attitudes, respondents being evaluated for clinical purposes are relatively likely to be motivated to reveal themselves honestly, whereas those being examined for administrative purposes are relatively likely to be intent on making a certain kind of impression. Respondents attempting to manage the impression they give tend to show themselves not as they are but as they think the person requesting the evaluation would view favorably. Typically such efforts at impression management take the form of denying one's limitations, minimizing one's shortcomings, attempting to put one's very best foot forward, and concealing whatever might be seen in a negative light.

Exceptions to this general trend are not uncommon, however. Whereas most persons being evaluated for administrative purposes want to make the best possible impression, some may be motivated in just the opposite direction. For example, a plaintiff claiming brain damage in a personal injury lawsuit may see benefit in making the worst possible impression on a neuropsychological examination.

## 10 Assessment Issues

Some persons being seen for clinical evaluations, despite having come of their own accord and recognizing that the assessment is being conducted for their benefit, may nevertheless be too anxious or embarrassed to reveal their difficulties fully. Whatever kind of impression respondents may want to make, the attitudes toward being examined that they bring with them to the assessment situation can be expected to influence the amount and kind of data they produce. These attitudes also have a bearing on the interpretation of assessment data. The further implications of impression management for malingering and defensiveness are discussed later in the chapter.

### Data Management Issues

A final set of considerations in collecting assessment information concerns appropriate ways of managing the data that are obtained. Examiners must be aware in particular of issues concerning (a) the use of computers in data collection; (b) the responsibility they have for safeguarding the security of their measures; and (c) their obligation, within limits, to maintain the confidentiality of what respondents report or reveal to them.

#### *Computerized Data Collection*

Software programs are available to facilitate the data collection process for most widely used assessment methods. Programs designed for use with self-report questionnaires typically provide for computerized administration of test items, automated coding of item responses to produce scale scores, and quantitative manipulation of these scale scores to yield summary scores and indices. For instruments that require examiner administration and coding (e.g., a Wechsler intelligence test), software programs accept test scores entered by the examiner and translate them into the test's quantitative indices (e.g., the Wechsler IQ and Index scores). Many of these programs store the test results in files that can later be accessed or exported, and some even provide computational packages that can generate descriptive statistics for sets of test records held in storage.

These features of computerized data management bring several benefits to the process of collecting assessment information. Computerized administration and coding of responses help respondents avoid mechanical errors in filling out test forms manually, and they eliminate errors that examiners sometimes make in scoring these responses. (See Allard & Faust, 2000.) For measures that require examiner coding and data entry, the utility of the results depends on accurate coding and entry, but, once these raw data are entered, software programs eliminate examiner error

in calculating summary scores and indices from them. The data storage features of many software programs facilitate assessment research, particularly for investigators seeking to combine databases from different sources, and they can also help examiners meet requirements in most states and many agencies for keeping assessment information on file for some period of time. Presumably for these reasons, many assessment psychologists have indicated in surveys that they use software for test scoring and feel comfortable doing so (McMinn, Ellens, & Soref, 1999).

Computerized collection of assessment information has some potential disadvantages as well, however. When assessment measures are administered by computer, first of all, there is a possibility that the reliability of the data collected could be compromised by a lack of equivalence between the automated procedure and the noncomputerized version of it. As elaborated by Butcher, Perry, and Dean (2009) and by Butcher, this volume, however, available data suggest fairly good equivalence for computerized administrations based on pencil-and-paper questionnaires, especially those used in personality assessment. (See also Butcher, Perry, & Hahn, 2004.) Specifically with respect to the MMPI, for example, a meta-analysis by Finger and Ones (1999) of research comparing computerized with booklet forms of the inventory showed them to be psychometrically equivalent. Good congruence with the original measures has yet to be demonstrated for computerized versions of structured clinical interviews, however, and Garb (2007) has reported only "poor to fair" concordance between diagnoses based on clinician interviews and diagnoses based on computer interviews.

A second potential disadvantage of computerized data collection derives from the ease with which it can be employed. Although frequently helpful to knowledgeable assessment professionals, automated procedures also simplify psychological testing for untrained and unqualified persons who lack assessment skills and would not be able to collect test data without the aid of a computer. The availability of software programs thus creates possibilities for assessment methods to be misused and respondents to be poorly served. Such outcomes are not an inescapable by-product of computerized assessment procedures, however. They constitute instead an abuse of technology by uninformed and irresponsible practitioners.

#### *Test Security*

*Test security* refers to restricting public access to test materials and answers to test items. Such restrictions address two important considerations in psychological assessment. First, publicly circulated information about

tests can undermine their validity, particularly if they comprise items with right and wrong or more or less preferable answers. Prior exposure to tests of this kind and information about correct or preferred answers can affect how persons respond to them and prevent an examiner from being able to collect a valid protocol. The validity of test findings is especially questionable when a respondent's prior exposure has included specific coaching in how to answer certain questions.

As for relatively unstructured assessment procedures that have no right or wrong answers, even on these measures various kinds of responses carry particular kinds of interpretive significance. Hence, the possibility also exists on relatively unstructured measures that a person can be helped to make a certain kind of impression by pretest instructions concerning what various types of responses are taken to signify. However, the extent to which public dissemination of information about the inferred meaning of responses does in fact compromise the validity of relatively unstructured measures has not been examined empirically and is a subject for further research.

Second, along with helping to preserve the validity of obtained results, keeping assessment measures secure protects test publishers against infringement of their rights by pirated or plagiarized copies of their products. Ethical assessors respect copyright law by not making or distributing copies of published tests, and they take appropriate steps to prevent test forms, test manuals, and assessment software from falling into the hands of persons who are not qualified to use them properly or who feel under no obligation to keep them secure. Both the Ethical Principles and Code of Conduct (APA, 2002, section 9.11) and the Standards for Educational and Psychological Testing (AERA et al., 1999, p. 117) address this professional responsibility in clear terms.

These considerations in safeguarding test security become particularly relevant in forensic cases should an attorney request to have a psychological examination observed by a third party. Such intrusions on traditional examination procedures pose a threat to the validity of the obtained data in two respects. First, there is no way to judge or measure the impact of an observer on how the test taker chooses to respond. Second, the normative standards that guide test interpretation are derived from data obtained in two-person examinations, and there are no comparison data available for examinations conducted in the presence of an observer. Validity aside, exposing test items to an observer can compromise test security in the same way as distributing test forms or manuals to persons who are under no obligation to keep them confidential.

### **Confidentiality**

A third and related aspect of appropriate data management pertains to maintaining the confidentiality of a respondent's assessment information. Like certain aspects of safeguarding test security, confidentiality is an ethical matter in assessment psychology, not a substantive one. The key considerations in maintaining the confidentiality of assessment information, as specified in the Ethical Principles and Code of Conduct (APA, 2002, Section 4) and elaborated by Kitchener (2000), involve:

- clarifying the nature and limits of confidentiality with clients and patients prior to undertaking an evaluation;
- communicating information about persons being evaluated only for appropriate scientific or professional purposes and only to an extent relevant to the purposes for which the evaluation was conducted;
- disclosing information only to persons designated by respondents or to other duly authorized persons or entities, except when otherwise permitted or required by law; and
- storing and preserving respondents' records in a secure fashion.

Like the matter of informed consent discussed previously, confidentiality is elaborated as an ethical issue in Rey-Casserly and Koocher, this volume.

### **INTERPRETING ASSESSMENT INFORMATION**

Following the collection of sufficient relevant data, the process of psychological assessment continues with a phase of evaluation in which these data are interpreted. The interpretation of assessment data consists of drawing inferences and forming impressions concerning what the findings reveal about a respondent's psychological characteristics. Accurate and adequately focused interpretations result in summary descriptions of psychological functioning that can then be utilized in the final phase of the assessment process as a foundation for formulating conclusions and recommendations that answer referral questions. Reaching this output phase requires consideration during the evaluation phase of the basis on which inferences are drawn and impressions are formed, the possible effects on the findings of malingering or defensiveness, and effective ways of integrating data from diverse sources.

### Basis of Inferences and Impressions

The interpretation of assessment data involves four sets of alternatives with respect to how assessors go about drawing inferences and forming impressions about what these data indicate. Interpretations can be based on either empirical or conceptual approaches to decision making; they can be guided either by statistically based decision rules or by clinical judgment; they can emphasize either nomothetic or idiographic characteristics of respondents; and they can include more or less reliance on computer-generated interpretive statements. Effective assessment usually involves informed selection among these alternatives and some tailoring of the emphasis given each of them to fit the particular context of the individual assessment situation.

#### *Empirical and Conceptual Guidelines*

The interpretation of assessment information can be approached in several ways. In what may be called an *intuitive* approach, assessment decisions stem from impressions that have no identifiable basis in the data. Instead, interpretations are justified by statements like “It’s just a feeling I have about her” or “I can’t say where I get it from, but I just know he’s that way.” In what may be called an *authoritative* approach, interpretations are based on the pronouncements of well-known or respected assessment psychologists, as in saying “These data mean what they mean because that’s what Dr. Authority says they mean.” The intuition of unusually empathic assessors and reliance on authority by well-read practitioners who choose their experts advisedly may on occasion yield accurate and useful impressions, but both approaches have serious shortcomings.

Unless intuitive assessors can identify specific features of the data that help them reach their conclusions, their diagnostic sensitivity cannot be taught to other professionals or translated into scientifically verifiable procedures. Unless authoritative assessors can explain in their own words the basis on which experts have reached the conclusions being cited, they are unlikely to impress others as being professionally knowledgeable themselves or as knowing what to think without being told what to think by someone else. Moreover, neither intuitive nor authoritative approaches to interpreting assessment information are likely to be as consistently reliable as approaches based on empirical and conceptual guidelines.

Empirical guidelines to decision making derive from the replicated results of methodologically sound research. When a specific assessment finding has repeatedly been

found to correlate highly with the presence of a particular psychological characteristic, it is empirically sound to infer the presence of that characteristic in a respondent who displays that assessment finding. Conceptual guidelines to decision making consist of psychological constructs that provide a logical bridge between assessment findings and the inferences drawn from them. For example, if subjectively felt distress contributes to a person’s remaining in and benefiting from psychotherapy (for which there is evidence; see Clarkin & Levy, 2004), then it is reasonable to expect that moderate but not excessive elevations on test indices of subjectively felt distress will increase the likelihood of a favorable outcome in psychotherapy (which they do; see Butcher & Perry, 2008; Weiner, 2004).

Both empirical and conceptual guidelines to interpretation bring distinct benefits to the assessment process. Empirical perspectives are valuable because they provide a foundation for achieving certainty in decision making. The adequacy of psychological assessment is enhanced by quantitative data concerning the normative distribution and other psychometric properties of measurements that reflect dimensions of psychological functioning. Lack of such data limits the confidence with which assessors can draw conclusions about the implications of their findings. Without being able to compare an individual’s test responses with normative expectations, for example, or without a basis for estimating false positive and false negative possibilities in the measures they have used, assessors can be only speculative in attaching interpretive significance to their findings. Similarly, the absence of externally validated cutting scores detracts considerably from the certainty with which assessors can translate test scores into qualitative distinctions, such as whether a person is mildly, moderately, or severely depressed.

Conceptual perspectives are valuable in the assessment process because they provide some explanation of why certain findings are likely to identify certain kinds of psychological characteristics or predict certain kinds of behavior. Having such explanations in hand offers assessors the pleasure of understanding not only how their measures work but also why they work as they do; these explanations help assessors focus their attention on aspects of their data that are relevant to the referral question to which they are responding; and they facilitate the communication of results in terms that address characteristics of the person being examined and not merely characteristics of the data obtained. As a further benefit of conceptual formulations of assessment findings, they foster hypotheses concerning previously unknown or unexplored linkages between assessment findings and dimensions of

psychological functioning, and by so doing they help to extend the frontiers of knowledge.

Empirical guidelines are thus necessary to the scientific foundations of assessment psychology, as a basis for certainty in decision making, but they are not sufficient to bring assessment to its full potential. Conceptual guidelines do not by themselves provide a reliable basis for drawing conclusions with certainty. However, by enriching the assessment process with explanatory hypotheses, they point the way to increased understanding of why people are as they are and behave as they do.

For the purposes they serve, then, both empirical and conceptual guidelines should figure prominently in the interpretation of assessment information. At times, concerns about preserving the scientific respectability of assessment have led to assertions that only empirical guidelines constitute an acceptable basis for decision making and that unvalidated conceptual guidelines have no place in scientific psychology. McFall and Treat (1999), for example, maintain that “the aim of clinical assessment is to gather data that allow us to reduce uncertainty concerning the probability of events” (p. 215). From their perspective, the information value of assessment data resides in scaled numerical values and conditional probabilities. By contrast, Hunsley and Mash (2010), in a more balanced perspective on evidence-based assessment, offer this observation: “Psychological assessment is, however, much more than just the data collected from psychometrically strong measures. At its heart, assessment is inherently a decision-making task in which the psychologist must iteratively formulate and test hypotheses” (p. 8).

Let it further be observed in this regard that the river of scientific discovery can flow through inferential reasoning that suggests truths long before they are confirmed by replicated research findings. Newton grasped the reason that apples fall from trees well in advance of experiments demonstrating the laws of gravity, Einstein conceived his theory of relativity with full confidence that empirical findings eventually would prove him correct, and neither man has been challenged with respect to his credentials as a scientist. Even though empirical guidelines are on average more likely to produce reliable conclusions than conceptual formulations, logical reasoning concerning the implications of clearly formulated concepts can also generate conclusions that serve useful purposes and stand the test of time. The process of arriving at conclusions in individual case assessment is accordingly likely to involve creative as well as confirmatory aspects of scientific thinking, and the utilization of assessment to generate hypotheses and fuel speculation may in the

course of scientific endeavor increase rather than decrease uncertainty by virtue of identifying new alternative possibilities to consider.

### *Statistical Rules and Clinical Judgment*

Empirical guidelines for decision making have customarily been operationalized by using statistical rules to arrive at conclusions concerning what assessment data signify. Statistical rules for interpreting assessment data comprise empirically derived formulas, or algorithms, that provide an objective, actuarial basis for deciding what these data indicate. When statistical rules are applied to the results of a psychological evaluation, the formula makes the decision concerning whether certain psychological characteristics are present (as in deciding whether a respondent has a particular trait or disorder) or whether certain kinds of actions are likely to ensue (as in predicting the likelihood of a respondent’s becoming suicidal or performing well in some occupation). Statistical rules have the advantage of ensuring that examiners applying a formula correctly to the same set of data will always arrive at the same conclusion concerning what these data mean. As a disadvantage, however, the breadth of the conclusions that can be based on statistical rules and their relevance to referral questions are limited by the composition of the database from which they have been derived.

For example, statistical rules may prove helpful in determining whether a student has a learning disability but provide little information about the nature of this student’s disability; they may predict the likelihood of a criminal defendant’s behaving violently but offer few clues to the kinds of situations most likely to trigger this person’s becoming violent; and they may help identify an individual’s suitability for one type of position in an organization but be mute with respect to his or her suitability for other types of positions in the organization. In each of these instances, moreover, a statistical rule derived from a group of people possessing certain demographic characteristics (e.g., age, gender, ethnicity, socioeconomic status) and having been evaluated in a particular setting may not generalize to persons with different demographic characteristics evaluated in other kinds of settings. Garb (2000) has noted in this regard that “statistical-prediction rules are of limited value because they have typically been based on limited information that has not been demonstrated to be optimal, and they have almost never been shown to be powerful” (p. 31).

In other words, the scope of statistical rules is restricted to findings pertaining to the particular kinds of persons, psychological characteristics, and circumstances that were

anticipated in building these rules. For many of the varied types of people seen in actual assessment practice, and for many of the complex and specifically focused referral questions raised about these people, statistical rules that can by themselves provide fully adequate answers may be in short supply.

As a further limitation of statistical rules, they share with all quantified assessment scales some unavoidable artificiality that accompanies translation of numerical scores into qualitative descriptive categories. On the Beck Depression Inventory (BDI; Beck, Steer, & Brown, 1996), for example, a score of 14 to 19 is taken to indicate “mild depression” and a score of 20 to 28 indicates “moderate depression.” Hence two people who have almost identical BDI scores, one with a 19 and the other with a 20, will be described differently by the statistical rule, one as mildly depressed and the other as moderately depressed.

Similarly, in measuring intelligence with the Wechsler Adult Intelligence Scale–IV (WAIS–IV; Wechsler, 2008), a Full Scale IQ score of 109 calls for describing a person’s intelligence as “average,” whereas a person with almost exactly the same level of intelligence and a Full Scale IQ of 110 falls in the “high average” range. According to the WAIS–IV formulas, a person with a Full Scale IQ of 91, like someone with a 109 IQ, would be labeled as having “average” intelligence; a person with a Full Scale IQ of 119, like someone with a 110 IQ, would be labeled as having “high average” intelligence. Assessors can minimize this problem by adding some further specificity to the WAIS–IV categories, as in labeling a 109 IQ as the high end of the average range and a 110 IQ as the low end of the high-average range. Although additional categorical descriptions for more narrowly defined score ranges can reduce the artificiality in the use of statistical rules, there are limits to how many quantitative data points on a scale can be assigned a distinctive qualitative designation.

Conceptual guidelines for decision making have been operationalized in terms of clinical judgment, which consists of the cumulative wisdom that practitioners acquire from their experience. Clinical guidelines may come to represent the shared beliefs of large numbers of practitioners, but they emerge initially as impressions formed by individual practitioners. In contrast to the objective and quantitative features of statistical rules, clinical judgments constitute a subjective and qualitative basis for arriving at conclusions. When clinical judgment is applied to assessment data, decisions are made by the practitioner, not by a formula. Clinical judgments concerning the interpretive significance of a set of assessment data are consequently less uniform than actuarial decisions and less

likely to be based on established fact. The applicability of clinical judgments, however, is infinite. Their breadth and relevance are limited not by any database but by the practitioner’s capacity to reason logically about possible relationships between psychological characteristics identified by the assessment data and psychological characteristics relevant to addressing referral questions, whatever their complexity and specificity.

The relative merit of statistical rules and clinical judgment in the assessment process has been the subject of considerable debate since this distinction was first formulated by Meehl (1954) in a classic book, *Clinical versus Statistical Prediction*. Contemporary publications of note concerning this important issue include contributions by Ægisdóttir et al. (2006), Garb (1998, 2005), Grove and Meehl (1996), Grove, Zald, Lebow, Snitz, and Nelson (2000), and Westen and Weinberger (2005). Much of the literature on this topic has consisted of assertions and rebuttals concerning whether statistical methods generally produce more accurate assessment results than clinical methods. In light of the strengths and weaknesses inherent in both statistical prediction and clinical judgment, as elaborated by Spengler in this volume, such debate serves little purpose and is regrettable when it leads to disparagement of either approach to interpreting assessment data.

As testimony to the utility of both approaches, it is important to note that the creation of good statistical rules for making assessment decisions typically begins with clinically informed selection of both (a) test items, structured interview questions, and other measure components to be used as predictor variables and (b) psychological conditions, behavioral tendencies, and other criterion variables to which the predictor variables are expected to relate. Empirical methods of scale construction and cross-validation are then employed to shape these selected predictor variables into valid actuarial measures of the relevant criterion variables.

Hence good statistical rules should almost always produce more accurate results than clinical judgment, because they encompass clinical wisdom plus the sharpening of this wisdom by replicated research findings. Clinical methods of assessment at their best depend on the impressions and judgment of individual practitioners, whereas statistical methods at their best constitute established fact that has been built on clinical wisdom. Relying only on clinical judgment in decision-making situations for which adequate actuarial guidelines are available disregards fundamental principles of evidence-based practice. Even the best judgment of the wisest practitioner can at times be clouded by inadvertent bias, insufficient consideration of

base rates, and other sources of influence discussed in the final section of this chapter and in Reynolds and Suzuki, this volume. When there is a reasonable choice, assessment decisions are more advisedly based on established fact than on clinical impressions.

The previously noted diversity of people and of the circumstances that lead to their being referred for an evaluation mean that assessment questions regularly arise for which there are no available statistical rules, and patterns of assessment data often resemble but do not quite match the parameters for which replicated research has demonstrated certain correlates. When statistical rules cannot fully answer questions being asked, what are assessors to do in the absence of fully validated data? Decisions could be deferred, on the grounds that sufficient factual basis for a decision is lacking, and recommendations could be delayed, pending greater certainty about what recommendations should be made. Alternatively, assessors in a situation of uncertainty can supplement whatever empirical guidelines they have at their disposal with logical reasoning and clinical wisdom to arrive at conclusions and recommendations that are more responsive, more timely, and at least a little more likely to be helpful than saying nothing at all.

As these observations indicate, statistical rules and clinical judgment can properly be regarded as complementary components of effective decision making rather than as competing and mutually exclusive alternatives. Each brings value to assessment psychology, and each has a respectable place in it. Geisinger and Carlson (2009) comment in this regard that “it is time to move beyond both purely judgmental, speculative interpretation of test results as well as extrapolations from the general population to specific cases that do not much resemble the remainder of the population” (p. 109).

Assessment practice should accordingly be subjected to and influenced by research studies, lest it lead down blind alleys and detract from the pursuit of knowledge and the delivery of responsible professional service. Concurrently, however, lack of unequivocal documentation should not deter assessment psychologists from employing procedures and reaching conclusions that in their judgment will assist in meeting the needs of those who seek their help. Meehl (1957) himself allowed that, in the absence of research findings to provide guidance, clinicians may have to “use our heads instead of the formula”; as for when we should use our heads instead of the formula, he said, “‘Always’ and ‘Never’ are equally unacceptable” (p. 268). Zeldow (2009) has more recently echoed Meehl by observing that “clinicians will always have to deal with

uncertainty, values, and uniqueness; they will always have to use their heads for certain purposes and to rely on clinical expertise, not only because the appropriate research has not yet been conducted, but also because that is the nature of clinical work” (p. 1).

### *Nomothetic and Idiographic Emphasis*

Empirical guidelines and statistical rules constitute a basically nomothetic approach to interpreting assessment information, whereas conceptual guidelines and clinical judgment underlie a basically idiographic approach. Nomothetic interpretations address ways in which people resemble other kinds of people and share various psychological characteristics with many of them. Hence these interpretations involve comparisons between the assessment findings for the person being examined and assessment findings typically obtained from groups of people with certain known characteristics, as in concluding that “this person’s responses show a pattern often seen in people who feel uncomfortable in social situations and are inclined to withdraw from them.” The manner in which nomothetic interpretations are derived and expressed is thus primarily quantitative in nature and may even include specifying the precise frequency with which an assessment finding occurs in particular groups of people.

Idiographic interpretations, by contrast, address ways in which people differ from most other people and show psychological characteristics that are fairly unique to them and their particular circumstances. These interpretations typically comprise statements that attribute person-specific meaning to assessment information on the basis of general notions of psychological processes, as in saying that “this person gives many indications of being a passive and dependent individual who is more comfortable being a follower than a leader and consequently may not function effectively in an executive position.” Deriving and expressing idiographic interpretations is thus a largely qualitative procedure in which examiners are guided by informed impressions rather than by quantitative empirical comparisons.

In the area of personality assessment, both nomothetic and idiographic approaches to interpretation have a long and distinguished tradition. Nomothetic perspectives in personality assessment derive from the work of Cattell (1946), for whom the essence of personality resided in traits or dimensions of functioning that all people share to some degree and on which they can be compared with each other. Idiographic perspectives in personality theory were first clearly articulated by Allport (1937), who conceived the essence of personality as residing in the

uniqueness and individuality of each person, independently of comparisons to other people.

Over the years, assessment psychologists have at times expressed different convictions concerning which of these two traditions should be emphasized in formulating interpretations. Practitioners typically concur with Groth-Marnat (2010) that data-oriented descriptions of people rarely address the unique problems they may be having and that the essence of psychological assessment is an attempt “to evaluate an individual in a problem situation so that the information derived from the assessment can somehow help with the problem” (p. 3). Writing from a research perspective, however, McFall and Townsend (1998) grant that practitioners must of necessity provide idiographic solutions to people’s problems but maintain that “nomothetic knowledge is a prerequisite to valid idiographic solutions” (p. 325) and that the clinical science of assessment should accordingly focus on nomothetic variables.

In light of what has already been said about statistical and clinical prediction, these nomothetic/idiographic differences of opinion can be tempered by recognizing that idiography can readily be managed in a scientific fashion (see Westen & Weinberger, 2004), and clinicians seeking solutions to idiographic problems can freely draw on whatever nomothetic guidelines will help them frame accurate and useful interpretations. In this vein, Stricker’s (1997, 2007) concept of the “local clinical scientist” incorporates a frame of reference and a set of values that can characterize both office practitioners and laboratory researchers: “The local clinical scientist brings the attitudes and knowledge base of the scientist to bear on the problems that must be addressed by the clinician in the consulting room” (Stricker & Trierweiler, 2006, p. 37).

Issues of definition aside, then, there seems little to be gained by debating whether people can be described better in terms of how they differ from other people or how they resemble them. In practice, an optimally informative and useful description of an individual’s psychological characteristics and functioning will address both the person’s resemblance to and differences from other people in similar circumstances about whom similar referral questions have been posed. Nomothetic and idiographic perspectives thus complement each other, and balanced attention to both promotes the fullest possible understanding of a person who is being evaluated.

### *Computer-Generated Interpretive Statements*

For most published tests, there are software programs that not only assist in the collection of assessment

data, as already discussed, but also generate interpretive statements that describe the test findings and present inferences based on them. Like computerized data collection, computer-based test interpretation (CBTI) brings some distinct advantages to the assessment process. By virtue of its automation, CBTI guarantees a thorough scan of the test data and thereby eliminates human error due to overlooking items of information in a test protocol. CBTI also ensures that a pattern of test data will always generate the same interpretive statement, uniformly and reliably, thus eliminating examiner variability and bias as potential sources of error. Additionally CBTI can facilitate the teaching and learning of assessment methods, by using computer-generated narratives as an exercise requiring the learner to identify the test variables that have given rise to particular interpretive statements.

The potential benefits of computerizing test interpretations, as well as some drawbacks of doing so, are elaborated in Butcher, this volume. (See also Butcher, Perry, & Dean, 2009.) Four limitations of CBTI have a particular bearing on the extent to which examiners should rely on computer-generated statements in formulating and expressing their impressions. First, although test software generates interpretive statements by means of quantitative algorithmic formulas, these computer programs are not entirely empirically based. Instead, they typically combine empirically validated correlates of test scores with clinical judgments about what various patterns of scores are likely to signify, and many algorithms involve beliefs as well as established fact concerning what these patterns mean. Different test programs, and even different programs for the same test, vary in the extent to which their interpretive statements are research based. Although CBTI generally increases the validity and utility of test interpretations, considerable research remains to be done to place computerized interpretation on a solid empirical basis. (See Garb, 2000.) In the meantime, computer-generated interpretations will embody at least some of the strengths and weaknesses of both statistical and clinical methods of decision making, and, as noted by Garb (2007), “it is recommended that computer assessment be combined with clinical judgment” (p. 207).

Second, the previously noted limitation of statistical rules with respect to designating quantitative score ranges with qualitative descriptors carries over into CBTI algorithms. Cutting points must be established, below which one kind or degree of descriptive statement is keyed and above which a different kind or degree of description will be generated. Consequently, as illustrated earlier for the BDI and WAIS, two people who score similarly on

some test index or scale may be described by a computer narrative in quite different terms with respect to psychological characteristics measured by this index or scale.

Third, despite often referring specifically to the person who was examined (i.e., using the terms *he*, *she*, or *this person*) and thus giving the appearance of being idiographic, computer-generated interpretations do not describe the individual who took the test. Instead, these interpretations describe test protocols; that is, they indicate what research findings or clinical wisdom say about people who show the same or similar test scores and patterns as those in the protocol being scanned by the computer. Hence computer narratives are basically nomothetic, and most of them phrase at least some of their interpretive statements in terms of normative comparisons or even the specific frequencies with which the respondent's test patterns occur in certain groups of people. However, because no two people are exactly alike, and no individual matches any comparison group perfectly, some computer-generated interpretive statements may not describe a test taker accurately.

For this reason, well-developed test software narratives include a caveat indicating that (a) the interpretive statements in the report describe groups of people, not necessarily the person who took the test; (b) misleading and erroneous statements may occur as a reflection of psychological characteristics or environmental circumstances unique to the person being examined and not widely shared within any normative group; and (c) other sources of information and the assessor's judgment are necessary to determine which of the statements in an interpretive narrative apply to the respondent and which do not.

Fourth, the availability of computer-generated interpretive statements raises questions concerning their proper utilization in the preparation of an assessment report. Ideally, assessors should draw on computer narratives for some assistance, as for example in being sure that they have taken account of all of the relevant data, in checking for discrepancies between their own impressions and the conclusions reached by the machine, and perhaps in getting some guidance on how best to organize and what to emphasize in their report. Less ideal is using CBTI not merely for supportive purposes but as a replacement for assessors being able and willing to generate their own interpretations of the data they have obtained. Most of the assessment psychologists responding to the previously mentioned McMinn et al. (1999) survey reported that they never use CBTI as their primary resource for case formulation and would question the ethicality of doing so.

Even among ethical assessors, however, CBTI can present some temptations, because many computerized

narratives present carefully crafted sentences and paragraphs that communicate clearly and lend themselves to being copied verbatim into a psychological report. Professional integrity would suggest that assessors relying on computer-generated conclusions should either express them in their own words or, if they are copying verbatim, should identify the copied material as a quotation and indicate its source. Regrettably, the previously mentioned software accessibility that allows untrained persons to collect and score test protocols by machine also makes it possible for them to print out narrative interpretations and reproduce them fully or in part as their report, without indicating their source. Aside from its questionable ethicality, the verbatim inclusion of computer-generated interpretations in assessment reports is likely to be a source of confusion and error because of the previously mentioned fact that these printouts are nomothetically rather than idiographically based and hence often include statements that are not applicable to the person being examined.

### **Malingering and Defensiveness**

Malingering and defensiveness consist of conscious and deliberate attempts by persons being examined to falsify the information they are giving and thereby mislead the examiner. *Malingering* involves intent to present oneself as being worse off psychologically than is actually the case and is commonly referred to as *faking bad*. *Defensiveness* involves seeking to convey an impression of being better off than one actually is and is commonly called *faking good*. Both faking bad and faking good can range in degree from slight exaggeration of problems and concerns, or of assets and capabilities, to total fabrication of difficulties never experienced or accomplishments never achieved. These two types of efforts to mislead examiners arise from different kinds of motivation, but usually both types can be detected from patterns of inconsistency that appear in the assessment data, unless respondents have been coached to avoid them.

### **Identifying Motivations to Mislead**

People who fake bad during psychological assessments are usually motivated by some specific reason for wanting to appear less capable or more disturbed than they really are. In clinical settings, for example, patients who are concerned about not getting as much help or attention as they would like to receive may exaggerate or fabricate symptoms in order to convince a mental health professional that they should be taken into psychotherapy, that they should be seen more frequently in outpatient treatment

than is currently the case, or that they should be admitted to an inpatient facility or kept in one if they are already there. In forensic settings, plaintiffs seeking damages in personal injury cases may mangle the extent of their neuropsychological or psychosocial impairments in hopes of increasing the amount of the settlement they receive, and defendants in criminal actions may mangle psychological disturbance in hopes of being able to minimize the penalties that are to be imposed on them. In employment settings, claimants may mangle inability to function in order to begin or continue receiving disability payments or unemployment insurance.

People who fake good during psychological assessments, in an effort to appear more capable or better adjusted than they really are, also show a variety of motivations related to the setting in which they are being evaluated. Defensive patients in clinical settings may try to conceal the extent of their difficulties when they hope to be discharged from a hospital to which they were involuntarily committed or when they would like to be told or have others told that they do not have any significant psychological problems for which they need treatment. In forensic settings, making the best possible impression can be a powerful inducement to faking good among divorced parents seeking custody of their children and among prison inmates requesting parole. In personnel settings, applicants for positions, candidates for promotion, and persons asking for reinstatement after having been found impaired have good reasons for putting their best foot forward during a psychological evaluation, even to the extent of overstating their assets and minimizing their limitations.

#### ***Detecting Malingering and Defensiveness***

Attempts to mislead psychological assessors usually produce patterns of inconsistency that provide dependable clues to malingering and defensiveness. In the case of efforts to fake bad, these inconsistencies are likely to appear in three different forms. First, malingerers often produce inconsistent data within individual assessment measures. Usually referred to as *intratest scatter*, this form of inconsistency involves failing relatively easy items on intelligence or ability tests while succeeding on more difficult items of the same kind or responding within the normal range on some portions of a personality test but in an extremely deviant manner on other portions of the same test.

A second form of inconsistency frequently found in the assessment data of malingerers occurs between test results and the examiner's behavioral observations. In some instances, for example, people who appear calm and relaxed

during an interview, talk clearly and sensibly about a variety of matters, and conduct themselves in a socially appropriate fashion produce test protocols similar to those seen in people who are extremely anxious or emotionally upset, incapable of thinking logically and coherently, out of touch with reality, and unable to participate comfortably in interpersonal relationships. Such discrepancies between test and interview data strongly suggest deceptive tactics intended to create a false impression of disturbance.

The third form of inconsistency that proves helpful in detecting malingering consists of a sharp discrepancy between the interview and test data collected by the examiner and the respondent's actual circumstances and past history as reported by collateral sources or recorded in formal documents. In these instances, the person being evaluated may talk and act strangely during an interview and give test responses strongly suggestive of serious psychological disturbance but never previously have seen a mental health professional, received counseling or psychotherapy, been prescribed psychotropic medication, or been considered by friends, relatives, teachers, or employers to have any emotional problems. Such contrasts between serious impairments or limitations suggested by the results of an examination and a life history containing little or no evidence of these impairments or limitations provide good reason to suspect malingering.

Defensiveness in an effort to look good is similarly likely to result in inconsistencies in the assessment data that help to detect it. Most common in this regard are guarded test protocols and minimally informative interview responses that fall far short of reflecting a documented history of psychological disorder or problem behavior. Although being guarded and tight-lipped may successfully conceal difficulties, it also alerts examiners that a respondent is not being forthcoming and that the data being obtained probably do not paint a full picture of his or her psychological problems and limitations. As another possibility, fake-good respondents may, instead of being guarded and closemouthed, become quite talkative and expansive in an effort to impress the examiner with their admirable qualities and many capabilities, in which case the assessment information becomes noteworthy for claims of knowledge, skills, virtues, and accomplishments that far exceed reasonable likelihood. These and other guidelines for the clinical detection of efforts to mislead assessors by faking either good or bad are elaborated by Rogers (2008a) and by Berry, Sollman, Schipper, Clark, and Shandera (2009).

Most self-report inventories include validity scales that are based on inconsistent and difficult-to-believe

responses that often can help to identify malingering and defensiveness. (See Greene, 2008.) A variety of specific interview, self-report, and ability measures have also been developed along these lines to assist in identifying malingering, including the Structured Interview of Reported Symptoms (SIRS; Rogers, Gillis, Dickens, & Bagby, 1991; Rogers, Jackson, Sewell, & Salekin, 2005) and the Test of Memory Malingering (TOMM; Teichner & Wagner, 2004; Tombaugh, 1997). (See also Berry & Schipper, 2008; Rogers, 2008b; Smith, 2008.)

Commonly used performance-based assessment instruments do not include formal validity scales but nevertheless are sensitive to inconsistencies in performance that suggest malingering or defensiveness. (See Ganellen, 2008; Sewell, 2008.) Moreover, because relatively unstructured measures convey less meaning to respondents than self-report questionnaires concerning what their responses might signify, there is reason to believe that they may be less susceptible to impression management or even that the fakability of an assessment instrument is directly related to its face validity (Bornstein, Rossner, Hill, & Stepanian, 1994; Weiner, 2005). This does not mean that unstructured measures like the Rorschach Inkblot Method and Thematic Apperception Test are impervious to malingering and defensiveness, which they are not, but only that efforts to mislead may be more obvious on these measures than on relatively structured measures and may be less likely to convey a desired impression.

### **Coaching**

A companion issue to the ease or difficulty of faking assessment measures is the extent to which respondents can be taught to deceive examiners with a good-looking or bad-looking performance. Research findings indicate that even psychologically naive participants who are given some information about the nature of people who have certain disorders or characteristics can shape their test behaviors to make themselves resemble a target group more closely than they could have done without such instruction. Misleading results are even more likely to occur when respondents are coached specifically in how to answer certain kinds of questions and avoid elevating validity scales (Gorny & Merten, 2007; Storm & Graham, 2000; Victor & Abeles, 2004). The group findings in these research studies have not yet indicated whether a generally instructed or specifically coached respondent can totally mislead an experienced examiner in actual practice, without generating any suspicion that the obtained results may not be valid, and this remains a subject for further investigation.

With further respect to individual assessments in actual practice, however, there are reports in the literature of instances in which attorneys have coached their clients in how to answer questions on self-report inventories (e.g., Lees-Haley, 1997; Wetter & Corrigan, 1995; Youngjohn, 1995), and a website on the Internet claims to provide a list of supposed good and bad responses for each of the 10 Rorschach inkblots. As previously mentioned in discussing test security, prior knowledge of test questions and answers can detract from the practical utility of psychological assessment methods that feature right and wrong answers. The confounding effect of pretest information on unstructured measures, for which correct or preferable answers are difficult to specify out of context, may be minimal, but the susceptibility of these measures to successful deception by well-coached respondents is another topic for future research. Less uncertain are the questionable ethics of persons who coach test takers in dishonesty and thereby thwart the legitimate purposes for which these respondents are being evaluated.

### **Integrating Data Sources**

As noted at the beginning of this chapter, psychological assessment information can be derived from administering tests, conducting interviews, observing behavior, speaking with collateral persons, and reviewing historical documents. Effective integration of data obtained from such multiple sources calls for procedures based on the previously described additive, confirmatory, and complementary functions served by a multimethod test battery. In some instances, for example, a respondent may during an interview report a problem for which there is no valid test index (e.g., having been sexually abused) and may demonstrate on testing a problem that is ordinarily not measured by interview data (e.g., poor perceptual-motor coordination). These two data sources can then be used additively to suggest that the person may have both a chronic traumatic stress disorder and a neuropsychological impairment. In another instance, a person who describes him- or herself during an interview as being a bright, well-educated individual with good leadership skills and a strong work ethic, and who then produces reliable documents attesting these same characteristics, offers assessors an opportunity for confirmatory use of these different data sources to lend certainty to a positive personnel report.

A third and somewhat more complicated set of circumstances may involve a respondent who behaves pleasantly and deferentially toward the assessor, reports being a kindly and even-tempered person, and produces limited

and mostly conventional test responses that fall in the normal range. At the same time, however, the respondent is described by friends and relatives as an angry and abusive person, and police reports show an arrest record for assault and domestic violence. Familiar to forensic psychologists consulting in the criminal justice system, this pattern of discrepant data usually can be explained by using them in a complementary fashion to infer defensiveness and a successful fake-good approach to the interviewing and testing situations. As a further example in educational settings, if a student whose poor grades suggest limited intelligence but whose test performance indicates an above-average IQ and no specific learning disability, assessors can draw in a complementary fashion on these divergent data to infer the possibility of psychologically determined underachievement.

Because of the increased understanding of people that can accrue from integrating multiple sources of information, thorough psychological evaluation utilizes all of the available data during the interpretation phase of the assessment process. This consideration in conducting psychological assessments touches on the question of how much data should be collected for the purposes of an evaluation. Theoretically, there can never be too much information in an assessment situation. There may be redundant information that provides more confirmatory evidence than is needed, and there may be irrelevant information that serves no additive function in answering the referral question, but examiners can choose to discard the former and ignore the latter. Moreover, all of the test, interview, and observational data that are collected reflect some psychological characteristics of people being evaluated and may therefore signify something important to know about them.

Practically speaking, however, there are limits to how much assessment information should be collected to guide the formulation of interpretations. Psychological assessors are responsible for conducting evaluations in a cost-effective manner that provides adequate responses to referral questions with the least possible expense of time and money. As noted previously, practitioners who provide and charge for services that they know will have little import are exploiting the recipients of their services and jeopardizing their own professional respectability. Assessment psychologists may differ in the amount and kind of data they regard as sufficient for a fully adequate evaluation, but they are ethically obligated to refrain from collecting information beyond what they genuinely believe could be helpful.

With respect to providing answers to referral questions, two additional guidelines can help assessment psychologists draw wisely and constructively on the assessment

data at their disposal. First, by taking account of indications of both psychological strengths and weaknesses in people they examine, assessors can present a balanced description of their assets and liabilities. Psychological assessment has tended over the years to address mainly what is wrong with people while giving insufficient attention to their adaptive capacities and admirable qualities. In keeping with a currently increasing focus in psychology on wellness, happiness, optimism, and other positive elements of the human condition, assessment serves its purposes best when the interpretive process gives full measure to coping strengths as well as functioning limitations. (See Duckworth, Steen, & Seligman, 2005; Linley, Harrington, & Garcea, 2010; Seligman & Csikszentmihalyi, 2000.)

Second, by recognizing that the strength of the evidence supporting the inferences and impressions they derive from assessment data is likely to vary, examiners can couch their interpretive statements in language that conveys their level of confidence in what they have to say. Most respondents provide clear and convincing evidence of at least some psychological characteristics, which examiners can then appropriately report in what may be called the *language of certainty*. The language of certainty states in direct terms what people are like and how they are likely to conduct themselves, as in saying “This student has a moderately severe reading disability,” or “Mr. A appears to be an impulsive person with limited self-control,” or “Ms. B is an outgoing and gregarious person who seeks out and enjoys interpersonal relationships.” For other characteristics of a person being evaluated, the evidence may be fragmentary or suggestive rather than compelling and conclusive, in which case impressions are properly reported in what may be called the *language of conjecture*. Conjectural language suggests or speculates about possible features of a person’s nature or behavioral tendencies, as in saying “There is some evidence to suggest that this child may have an auditory processing deficit,” or “She appears prone on occasion to jump to conclusions and exercise poor judgment, which might limit her ability to make good decisions,” or “The data provide some basis for speculating that his lack of effort represents a passive-aggressive way of dealing with underlying anger and resentment he feels toward people who have made a lot of demands on him.”

## USING ASSESSMENT INFORMATION

The assessment process culminates in using descriptions of psychological characteristics and behavioral tendencies to formulate conclusions and recommendations.

Interpretations of assessment information are now translated into their implications for various decisions, and the overall purpose and eventual goal of assessment can be conceived as a way of facilitating decision making about classification, selection, placement, diagnosis, and treatment of people being evaluated. In this output phase, however, account must be taken of the fact that assessment data and the descriptions they generate may have different implications for different kinds of people living in different circumstances. Most important in this regard are possible sources of bias, applicable base rates, value judgments that call for cutting-score adjustments, and the cultural background and social context of the person being evaluated. Good assessment decisions depend on recognizing these considerations and preventing them from exerting undue influence on conclusions and recommendations.

### Bias and Base Rates

As elaborated in Reynolds and Suzuki, this volume, bias occurs in the utilization of assessment information when examiners allow preconceived notions and previously held beliefs to influence how they view the implications of their data. Assessment bias may arise either inadvertently, from attitudes of which examiners are unaware, or consciously, as purposeful intent on their part. Whether inadvertent or intentional, assessment bias takes the form of expectations that derive from demographic beliefs, environmental impressions, or epidemiological notions and that affect the meaning assigned to a set of findings.

As an example of demographic beliefs, an assessor who thinks that older people and males are generally likely to perform better as managers than younger people and females may advise hiring a 45-year-old man and not hiring a 30-year-old woman for a managerial position, even if their psychological assessment information would be seen by most examiners as comparable or even favoring the female candidate. Similarly, assessors who harbor a conviction that blue-collar African Americans are generally less likely to respond to psychotherapy than white-collar Caucasians may discourage psychotherapy for the former and recommend it for the latter, even when they are looking at assessment information showing equivalent treatment accessibility.

Environmental impressions as a source of biased expectations relate to the setting in which assessors are conducting an evaluation. Psychologists working in an inpatient facility in which a large percentage of patients are psychotically disturbed come to expect most of the people they examine to be psychotic, at least on

admission, and accordingly they may be inclined to infer psychosis from a set of assessment data that would not have led them to this conclusion had they obtained it in an outpatient clinic in which psychotic disturbance is rarely seen. Similarly, psychologists assessing prison inmates, among whom antisocial personality disorder is commonly found, may be more likely to expect and diagnose this disorder than they would if they were working with similar data in a university counseling center.

As for epidemiological notions, examiners may be consciously or inadvertently influenced in the conclusions they draw by how they view the nature and incidence of various conditions. Those who believe that borderline personality disorder is widespread are likely to diagnose this condition more frequently than those who think this diagnostic category lacks precision and is too frequently used. Those who believe that attention-deficit/hyperactivity disorder (ADHD) occurs mainly in boys and adolescent anorexia mainly in girls are relatively unlikely to diagnose ADHD in girls and anorexia in boys.

In all such instances of possible untoward influence stemming from demographic, environmental, and epidemiological expectations, the challenge for assessment psychologists is to recognize their personal biases and prevent them as much as possible from affecting the conclusions they derive from their assessment data and the recommendations they base on these conclusions. As indicated by the examples just given, however, epidemiological and environmental expectations may have some basis in fact. There *are* more psychotic patients in hospital than in clinic populations, there *are* more antisocial individuals in prison than on college campuses, and there *are* gender differences in the incidence of ADHD and anorexia. From a strictly actuarial point of view, then, being hospitalized does increase the probability of being psychotic, being incarcerated does increase the probability of being antisocial, and being male or female does increase the probability of being attention disordered or anorexic, respectively. To take adequate account of such actual setting and group differences while preventing them from resulting in biased conclusions, examiners need to be alert to whatever base-rate information may be applicable in the individual case.

*Base-rate information* consists of the expected frequency of a characteristic or behavior in a particular set of persons or circumstances. Attention to applicable base rates provides a way of estimating the utility of assessment procedures, particularly with respect to their efficiency in assessing rare events. As first identified by Meehl and Rosen (1955), base rates can become problematic for

measuring instruments when the expected frequency of an event falls very far below 50%. For example, in a clinical setting in which 10% of the patients are suicidal, a valid test of suicidality that has a hit rate of 60% (i.e., is correct 60% of the time in identifying people in general as suicidal or nonsuicidal) is technically less efficient than simply calling all of the patients nonsuicidal, which would be correct 90% of the time.

Although technically correct from a psychometric perspective, this type of base-rate emphasis on efficiency does not always satisfy priorities in actual assessment practice. Assessment methods that are inefficient in assessing suicidality, given its low base rate even in most patient populations, may nevertheless correctly identify a subgroup of patients in whom suicidal behavior is relatively likely to occur. An examiner can then use this information to recommend suicide precautions for this apparently suicidal subgroup, which is preferable to overlooking the self-destructive potential of the high-risk group by exercising the technically more efficient option of calling all of the patients nonsuicidal.

The base-rate problem can also be minimized by focusing assessment efforts on restricted populations in which the expected frequency of the characteristic being evaluated is less rare than in the general population. Kamphuis and Finn (2002) note in this regard that the more closely a base rate approximates 50%, the better prospects a valid measure has of improving on the efficiency of concluding that either everyone or no one has a certain characteristic or behavioral tendency. As an example of increasing the base rate by restricting the population, efficient prediction of violent behavior among people in general is difficult to achieve, because most people are nonviolent. In a population of criminal offenders, however, many of whom have a history of violence, a valid measure of violence potential may prove quite efficient in identifying those at highest risk for violent behavior in the future.

### Value Judgments and Cutting Scores

Value judgments in the present context relate to the purposes for which a respondent is being evaluated and the frequency of false-positive and false-negative outcomes that an assessment variable is likely to produce. False-positive outcomes result in concluding that people have certain conditions and tendencies that they in fact do not have; false-negative outcomes result in concluding that people lack certain conditions and tendencies that in fact do characterize them. When assessments are being conducted to assist in making decisions about psychological

characteristics and behavioral tendencies that most people would consider as undesirable, such as being suicidal or homicidal, false positives may be of less concern than false negatives. A false-positive decision concerning dangerousness might result in a person's being unnecessarily supervised or even restrained, which is a regrettable but not a fatal outcome. A false-negative decision, however, by failing to identify existing dangerousness to oneself or others, can result in loss of life.

Conversely, false-positive outcomes may be more problematic than false-negative outcomes when referral questions concern desirable characteristics and probable outcomes, such as whether a person should be given a scholarship, a job, a promotion, or a parole. False negatives in this kind of assessment situation may result in denying people opportunities for which they are qualified and deserving, which is disadvantageous and perhaps unfair to them as individuals. However, when false positives lead to promoting personnel to positions of responsibility that exceed their capacities or paroling felons whose criminal tendencies have not abated, many people other than the individual are likely to suffer regrettable consequences.

In relation to such value judgments, then, a set of assessment data may have different implications in different assessment circumstances and thereby call for assessors to select carefully the cutting scores they use in formulating their conclusions and recommendations. For quantifiable dimensions of assessment that correlate positively with the presence of a characteristic or behavioral tendency, moving up the numerical scale produces a progressively decreasing percentage of false positives and moving down the scale produces a progressively decreasing percentage of false negatives; just the opposite is the case for assessment dimensions that are inversely correlated with what they measure. As a way of deciding the implications of assessment findings in a particular circumstance, cutting scores can thus be selected to minimize the likelihood of false-positive outcomes in examinations concerned with desirable consequences and to minimize false-negative outcomes in the estimation of undesirable consequences.

### Culture and Context

Just as assessment information may have different implications for people examined in different settings and for different purposes, it may also vary in its significance for respondents coming from different cultures or living in different social contexts. Hence the utilization phase of the assessment process must always take account of how characteristics of individuals identified in the interpretive

phase are likely to affect their psychological functioning in their particular circumstances. Attention to cross-cultural influences has a long history in assessment psychology (see, e.g., Hallowell, 1951; Lindzey, 1961) and has seen a recent resurgence of interest, as described in Geisinger and McCormick, this volume, and in contributions by Dana (2000, 2005), Hambleton, Merenda, and Spielberger (2005), and Suzuki and Ponterotto (2008).

The distinction drawn in this overview of the assessment process between interpreting and using assessment information provides guidelines for a two-step process in taking account of background and situational differences among respondents. The interpretive phase of assessment provides the first step, which consists of arriving at descriptive statements that identify a respondent's psychological characteristics as they exist independently of his or her cultural context and circumstances. Having superior intelligence, being orderly and compulsive, experiencing memory loss, being emotionally reserved, having an assertive and competitive bent, and being prone to acute anxiety in unfamiliar surroundings are examples of characteristics that define the nature of the individual. Any such characteristics identified by the assessment data obtained from people will be present in these individuals regardless of where they live, from whence they come, and in what they are involved.

The utilization phase of the assessment process provides the second step, which consists of being sufficiently sensitive to respondents' cultural and experiential contexts to estimate accurately the implications of their psychological characteristics in their particular life circumstances. Especially important to consider in this regard are whether the psychological characteristics people show are likely to prove adaptive or maladaptive in their everyday world and what kinds of successful or unsuccessful adaptations might result from these characteristics.

Research findings document that cultural differences can lead to cross-cultural variation in modal psychological characteristics and that the demands and expectations people face often determine the implications and consequences of certain characteristics, especially with respect to how adaptive these characteristics are. (See Kitayama & Cohen, 2007; Stevens & Gielen, 2007.) For example, a generally passive, dependent, agreeable, and acquiescent person may experience adjustment difficulties in a cultural context that values autonomy, self-reliance, assertiveness, and competitiveness. Conversely, a fiercely independent and highly competitive person might feel comfortable and flourish psychologically in a subculture that values individuality and assertiveness but feel alienated and adapt poorly

in a society that subordinates individual needs and preferences to the wishes and welfare of the group and in which a passive and acquiescent person could get along very well.

These contextual influences on the implications of psychological characteristics extend to specific circumstances in persons' lives as well as their broader sociocultural contexts. A modest level of intelligence can be a source of comfort and success to young people whose personal and family expectations are that they graduate from high school but a source of failure and dismay to those for whom graduation from a prestigious college is a minimum expectation. Similarly, a person with good coping skills and abundant adaptive capacities who is carrying a heavy burden of responsibilities and confronting numerous obstacles to meeting them may be susceptible to anxiety, irritability, and other consequences of a stress overload, whereas a person with limited coping skills and few adaptive capacities who is leading a narrowly restricted life and facing very few demands may be able to maintain a comfortable psychological equilibrium and experience little in the way of subjectively felt distress. A contemplative person who values being painstakingly careful in completing tasks and reaching conclusions may perform well in a job situation that calls for accuracy and thoroughness and involves little time pressure; this same person may perform poorly in a position that imposes strict deadlines or requires rapid decisions on the basis of sketchy information and in which a more decisive and action-oriented person would function effectively.

As illustrated by this final example and those that have preceded it in this chapter, psychological assessment is a complex process. Diverse perspectives and attention to interacting variables are necessary in assessment psychology as elsewhere in behavioral science to expand knowledge and guide its practical application, and there is little to be gained from doctrinaire pronouncements or unidimensional approaches. Collecting, interpreting, and using assessment data effectively call for giving just due to each of the issues identified in this introduction to the assessment process, and the 24 chapters that follow are designed for this purpose.

## REFERENCES

- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., . . . Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical judgment. *Counseling Psychologist, 34*, 341–382.
- Allard, G., & Faust, D. (2000). Errors in scoring objective personality tests. *Assessment, 7*, 119–129.

- Allen, J., & Walsh, J. A. (2000). A construct-based approach to equivalence: Methodologies for cross-cultural/multicultural personality assessment research. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 63–86). Mahwah, NJ: Erlbaum.
- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York, NY: Holt.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, *57*, 1060–1073.
- Beck, A. T., Steer, R. A., & Brown, T. K. (1996). *Beck depression inventory manual* (2nd ed.). San Antonio, TX: Psychological Corporation.
- Berry, D. T. R., & Schipper, L. J. (2008). Assessment of feigned cognitive impairment using standard neuropsychological tests. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 232–252). New York, NY: Guilford Press.
- Berry, D. T. R., Sollman, S. J., Schipper, L. J., Clark, J. A., & Shandera, A. L. (2009). Assessment of feigned psychological symptoms. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (pp. 613–642). New York, NY: Oxford University Press.
- Bornstein, R. F., Rossner, S. C., Hill, E. L., & Stepanian, M. L. (1994). Face validity and fakability of objective and projective measures of dependency. *Journal of Personality Assessment*, *63*, 363–386.
- Butcher, J. N., & Perry, J. N. (2008). *Personality assessment in treatment planning: Use of the MMPI-2 and BTPI*. New York, NY: Oxford University Press.
- Butcher, J. N., Perry, J. N., & Dean, B. E. (2009). Computer-based assessment. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (pp. 163–182). New York, NY: Oxford University Press.
- Butcher, J. N., Perry, J. N., & Hahn, J. (2004). Computers in clinical assessment: Historical developments, present status, and future challenges. *Journal of Clinical Psychology*, *60*, 331–345.
- Butcher, J. N., & Rouse, S. V. (1996). Personality: Individual differences and clinical assessment. *Annual Review of Psychology*, *47*, 87–111.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.
- Clarkin, J. F., & Levy, K. N. (2004). The influence of client variables on psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 194–226). Hoboken, NJ: Wiley.
- Cattell, R. B. (1946). *Description and measurement of personality*. New York, NY: World Book.
- Dana, R. H. (2005). *Multicultural assessment: Principle, applications, and examples*. Mahwah, NJ: Erlbaum.
- Dana, R. H. (Ed.). (2000). *Handbook of cross-cultural and multicultural personality assessment*. Mahwah, NJ: Erlbaum.
- Duckworth, A. L., Steen, T. A., & Seligman, M. E. P. (2005). Positive psychology in clinical practice. *Annual Review of Clinical Psychology*, *1*, 629–651.
- Exner, J. E., Jr. (2003). *The Rorschach: A comprehensive system. Vol. 1. Foundations* (4th ed.). Hoboken, NJ: Wiley.
- Finger, M. S., & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment*, *11*, 58–66.
- Finn, S. E. (1996). Assessment feedback integrating MMPI-2 and Rorschach findings. *Journal of Personality Assessment*, *67*, 543–557.
- Ganellen, R. J. (1996). *Integrating the Rorschach and the MMPI in personality assessment*. Mahwah, NJ: Erlbaum.
- Ganellen, R. J. (2008). Rorschach assessment of malingering and defense response sets. In C. B. Gacono & F. B. Evans (Eds.), *Handbook of forensic Rorschach assessment* (pp. 89–120). New York, NY: Routledge.
- Garb, H. N. (1998). *Studying the clinician*. Washington, DC: American Psychological Association.
- Garb, H. N. (2000). Computers will become increasingly important for psychological assessment: Not that there's anything wrong with that! *Psychological Assessment*, *12*, 31–39.
- Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology*, *1*, 67–89.
- Garb, H. N. (2007). Computer-administered interviews and rating scales. *Psychological Assessment*, *19*, 4–12.
- Geisinger, K. F., & Carlson, J. F. (2009). Standards and standardization. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (pp. 99–111). New York, NY: Oxford University Press.
- Geisinger, K. F., Spies, R. A., Carlson, J. F., & Plake, B. S. (Eds.) (2007). *The seventeenth Mental Measurements Yearbook*. Lincoln: University of Nebraska Press.
- Gorny, I., & Merten, T. (2007). Symptom information—warning—coaching: How do they affect successful feigning in neuropsychological assessment? *Journal of Forensic Neuropsychology*, *4*, 71–97.
- Graham, J. R. (2006). *MMPI-2: Assessing personality and psychopathology* (4th ed.). New York, NY: Oxford University Press.
- Greene, R. L. (2008). Malingering and defensiveness on the MMPI. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 159–181). New York, NY: Guilford Press.
- Greene, R. L. (2011). *The MMPI-2/MMPI-2-RF: An interpretive manual* (3rd ed.). Boston: Allyn & Bacon.
- Groth-Marnat, G. (2010). *Handbook of psychological assessment* (5th ed.). Hoboken, NJ: Wiley.
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293–323.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19–30.
- Hallowell, A. I. (1951). The use of projective techniques in the study of the socio-psychological aspects of acculturation. *Journal of Projective Techniques*, *15*, 27–44.
- Hambleton, R. K., Merenda, P., & Spielberger, C. D. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Erlbaum.
- Heilbrun, K. (2001). *Principles of forensic mental health assessment*. New York, NY: Kluwer Academic/Plenum.
- Heilbrun, K., Grisso, T., & Goldstein, A. M. (2009). *Foundations of forensic mental health assessment*. New York, NY: Oxford University Press.
- Hogan, T. P. (2005). 50 widely used psychological tests. In G. P. Koocher, J. C. Norcross, & S. S. Hill III. *Psychologists' desk reference* (2nd ed., pp. 101–104). New York, NY: Oxford University Press.
- Hunsley, J., & Mash, E. J. (2010). The role of assessment in evidence-based practice. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (2nd ed., pp. 3–22). New York, NY: Guilford Press.
- Kamphuis, J. H., & Finn, S. E. (2002). Incorporating base rate information in daily clinical decision making. In J. N. Butcher (Ed.), *Clinical personality assessment* (2nd ed., pp. 257–268). New York, NY: Oxford University Press.
- Kitayama, S., & Cohen, D. (Eds.). (2007). *Handbook of cultural psychology*. New York, NY: Guilford Press.
- Kitchener, K. S. (2000). *Foundations of ethical practice, research, and teaching in psychology*. Mahwah, NJ: Erlbaum.
- Koocher, G. P., & Keith-Spiegel, P. (2008). *Ethics in psychology and the mental health professions: Standards and cases* (3rd ed.). New York, NY: Oxford University Press.

- Krishnamurthy, R., & Yalof, J. A. (2010). The assessment competency. In M. B. Kenkel & R. L. Peterson (Eds.), *Competency-based education for professional psychology* (pp. 87–104). Washington, DC: American Psychological Association.
- Kubiszyn, T. W., Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., . . . Eisman, E. J. (2000). Empirical support for psychological assessment in clinical health care settings. *Professional Psychology, 31*, 119–130.
- Lees-Haley, P. R. (1997). Attorneys influence expert evidence in forensic psychological and neuropsychological cases. *Assessment, 4*, 321–324.
- Lindzey, G. (1961). *Projective techniques and cross-cultural research*. New York, NY: Appleton-Century-Crofts.
- Linley, P. A., Harrington, S., & Garcea, N. (Eds.) (2010). *Oxford handbook of positive psychology*. New York, NY: Oxford University Press.
- Masling, J. M. (1966). Role-related behavior of the subject and psychologist and its effect upon psychological data. In D. Levine (Ed.), *Nebraska symposium on motivation* (pp. 67–104). Lincoln: University of Nebraska Press.
- Masling, J. M. (1998). Interpersonal and actuarial dimensions of projective testing. In L. Handler & M. J. Hilsenroth (Eds.), *Teaching and learning personality assessment* (pp. 119–135). Mahwah, NJ: Erlbaum.
- Matarazzo, J. D. (1990). Psychological assessment versus psychological testing. *American Psychologist, 45*, 999–1017.
- McFall, R. M., & Townsend, J. T. (1998). Foundations of psychological assessment: Implications for cognitive assessment in clinical science. *Psychological Assessment, 10*, 316–330.
- McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50*, 215–241.
- McMinn, M. F., Ellens, B. M., & Soref, E. (1999). Ethical perspectives and practice behaviors involving computer-based test interpretations. *Assessment, 6*, 71–77.
- Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.
- Meehl, P. E. (1957). When shall we use our heads instead of the formula. *Journal of Counseling Psychology, 4*, 268–273.
- Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194–216.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56*, 128–165.
- Millon, T. (2006). *Millon Clinical Multiaxial Inventory—III (MCMI—III) manual* (3rd ed.). Minneapolis, MN: Pearson Assessments.
- Monahan, J. (Ed.). (1980). *Who is the client?* Washington, DC: American Psychological Association.
- Morey, L. C. (2007). *Personality Assessment Inventory professional manual* (2nd ed.). Lutz, FL: Psychological Assessment Resources.
- Rogers, R. (2008a). Detection strategies for malingering and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 14–35). New York, NY: Guilford Press.
- Rogers, R. (2008b). Structured interviews and dissimulation. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 301–322). New York, NY: Guilford Press.
- Rogers, R., Gillis, J. R., Dickens, S. E., & Bagby, R. M. (1991). Standardized assessment of malingering: Validation of the Structure Inventory of Reported Symptoms. *Psychological Assessment, 3*, 89–96.
- Rogers, R., Jackson, R. L., Sewell, K. W., & Salekin, K. L. (2005). Detection strategies for malingering: A confirmatory factor analysis of the SIRS. *Criminal Justice and Behavior, 32*, 511–525.
- Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5–14.
- Sewell, K. W. (2008). Dissimulation on projective measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 207–217). New York, NY: Guilford Press.
- Smith, G. P. (2008). Brief screening measures for the detection of feigned psychopathology. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 323–339). New York, NY: Guilford Press.
- Stevens, M. J., & Gielen, U. P. (Eds.) (2007). *Toward a global psychology: Theory, research, intervention, and pedagogy*. Mahwah, NJ: Erlbaum.
- Storm, J., & Graham, J. R. (2000). Detection of coached malingering on the MMPI-2. *Psychological Assessment, 12*, 158–165.
- Stricker, G. (1997). Are science and practice commensurable? *American Psychologist, 52*, 442–448.
- Stricker, G. (2007). The local clinical scientist. In S. G. Hofmann & J. Weinberger (Eds.), *The art and science of psychotherapy* (pp. 85–99). New York, NY: Routledge.
- Stricker, G., & Trierweiler, S. J. (2006). The local clinical scientist: A bridge between science and practice. *Training and Education in Professional Psychology, 5*, 37–46.
- Suzuki, L. A., & Ponterotto, J. G. (Eds.) (2008). *Handbook of multicultural assessment*. San Francisco, CA: Jossey-Bass.
- Teichner, G., & Wagner, M. T. (2004). The Test of Memory Malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology, 19*, 455–465.
- Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9*, 260–268.
- Turner, S. W., DeMers, S. T., Fox, H. R., & Reed, G. M. (2008). APA's guidelines for test user qualifications: An executive summary. In D. N. Bersoff (Ed.), *Ethical conflicts in psychology* (4th ed., pp. 279–282). Washington, DC: American Psychological Association.
- Victor, T. L., & Abeles, N. (2004). Coaching clients to take psychological and neuropsychological tests: A clash of clinical obligations. *Professional Psychology: Research and Practice, 35*, 373–379.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). Minneapolis, MN: Pearson.
- Weiner, I. B. (1989). On competence and ethicality in psychodiagnostic assessment. *Journal of Personality Assessment, 53*, 827–831.
- Weiner, I. B. (2003). *Principles of Rorschach interpretation* (2nd ed.). Mahwah, NJ: Erlbaum.
- Weiner, I. B. (2004). Rorschach Inkblot Method. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome evaluation* (Vol. 3, pp. 553–588). Mahwah, NJ: Erlbaum.
- Weiner, I. B. (2005). Integrative personality assessment with self-report and performance-based measures. In S. Strack (Ed.), *Handbook of personology and psychopathology* (pp. 317–331). Hoboken, NJ: Wiley.
- Weiner, I. B. (2009). Anticipating ethical and legal challenges in personality assessments. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (pp. 599–612). New York, NY: Oxford University Press.
- Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist, 59*, 595–613.
- Westen, D., & Weinberger, J. (2005). In praise of clinical judgment: Meehl's forgotten legacy. *Journal of Clinical Psychology, 61*, 1257–1276.
- Wetter, M. W., & Corrigan, S. K. (1995). Providing information to clients about psychological tests: A survey of attorneys' and law students' attitudes. *Professional Psychology, 26*, 474–477.
- Youngjohn, J. R. (1995). Confirmed attorney coaching prior to neuropsychological evaluation. *Assessment, 2*, 279–284.
- Zeldow, P. B. (2009). In defense of clinical judgment, credentialed clinicians, and reflective practice. *Psychotherapy: Theory, Research, Practice, Training, 46*, 1–10.