Report of the Task Force on

Test User Qualifications

Practice and Science Directorates

American Psychological Association

Approved by the APA Council of Representatives

August, 2000

Task Force Members

Stephen T. DeMers, EdD (Cochair)
Samuel M. Turner, PhD (Cochair)
Marcia Andberg, PhD
William Foote, PhD
Leaetta Hough, PhD
Robert Ivnik, PhD
Scott Meier, PhD
Kevin Moreland, PhD (deceased)
Celiane M. Rey-Casserly, PhD

Table of Contents

Preface

In response to an increasing number of requests from members and the public for

guidance on the qualifications that the American Psychological Association (APA) considers

important for test use, the APA Council of Representatives convened a Task Force on Test User

Qualifications in August 1996. The Board of Scientific Affairs appointed Samuel M. Turner,

PhD, as co-chair, and the Board of Professional Affairs and the Committee for the Advancement

of Professional Practice appointed Stephen DeMers, EdD, as the other co-chair. An additional

seven members were appointed by an extended consultative process and represent the following

areas of expertise specified by the Board and Council: clinical, industrial/organizational, school,

counseling, educational, forensic, and neuropsychology.

The Task Force met seven times between 1996 and 1999. Between and after these

meetings, drafts of the report were circulated, revised, and revised again. At various stages,

drafts of the document were reviewed by the governing bodies of APA, division associations,

state associations, APA members, and several outside organizations whose members use tests.

The members of the Task Force would like to thank the numerous psychologists and other test

users who reviewed and commented on earlier versions of this report. Many of their helpful

responses were incorporated in this final version, and we are grateful for their assistance. In

particular, the task force acknowledges the comments of APA members Wayne Camara, PhD;

Rodney Lowman, PhD; Kathleen O'Brien, PhD, Nancy T. Tippins, Ph.D, and Mary V. McGuire,

PhD, JD.

Support for the project was provided by the staff of the Practice and Science Directorates

of APA. In particular, the Task Force would like to thank Heather Roberts Fox, PhD, Geoffrey

I. Introduction

At the direction of the Council of Representatives of the American Psychological

Association (APA), the Task Force on Test User Qualifications (TFTUQ) was established in

October 1996. The goal of the task force was to develop guidelines that inform test users and the

general public of the qualifications that the APA considers important for the competent and

responsible use of psychological tests. The phrase *test user qualifications* refers to the

combination of knowledge, skills, abilities, training, experience, and, where appropriate,

credentials that the APA considers optimal for psychological test use. The guidelines in this

report are intended to apply to persons who use psychological tests in a variety of settings and

for diverse purposes. This report describes test user qualifications that the APA believes will best

serve the public.

The TFTUQ was established in part because of evidence that some current users of

psychological tests may not possess the knowledge and skill that the APA considers desirable for

optimal test use (e.g., Aiken, West, Sechrest, & Reno, 1990). Thus, it is hoped that these

guidelines will encourage training programs to make curricular changes that provide future test

users with a strong background in measurement theory and psychometrics, along with improved

skill in the administration, interpretation, and communication of test results. In addition, these

guidelines should encourage groups or individuals to obtain continuing education to improve

their use of psychological tests. The APA's goal in promulgating these guidelines is to encourage

the development of the knowledge, skills, abilities, and experiences that promote optimal testing

practices for the purpose of maintaining high standards in professional test use with the public.

Definition of Key Terms

Critical terms used in this document are defined as follows:

Psychological test: a measurement procedure for assessing psychological characteristics in which a sample of an examinee's behavior is obtained and subsequently evaluated and scored using a standardized process.

Test user: the person or persons responsible for the selection, administration, and scoring of tests; for the analysis, interpretation, and communication of test results; and for any decisions or actions that are based, in part, on test scores. Generally, individuals who simply administer tests, score tests and communicate simple or "canned" test results are not *test users*.

Test user qualifications: knowledge, skills, abilities, training, experience, and, where appropriate, credentials important for optimal use of psychological tests.

Assessment: a process that integrates test information with information from other sources; a process for evaluating behavior, psychological constructs, and/or characteristics of individuals or groups for the purpose of making decisions regarding classification, selection, placement, diagnosis, or intervention.

Context: the situation, purpose, or setting in which a test is being used; circumstances that determine when testing is appropriate for a person or group.

Scoring:　　　　　　application of test-specific rules to the responses or behavior of the test

taker to produce quantitative or qualitative data about the test taker or a

group of test takers.

Interpretation:　　　application of scientific knowledge and professional judgment to test data

to describe and/or make inferences about individual or group

characteristics or behavior.

Communication of test results: oral or written description and explanation of test findings to

others.

Supervision:　　　　the process of overseeing, directing, and assuming responsibility for the

actions of others involved in the testing process.

　　　　Scope of the Guidelines

　　　　The APA's purpose in developing these guidelines is to inform test users as well as

individuals involved with training programs, regulatory and credentialing bodies, and the public

about the qualifications that the APA considers important for the *optimal* use of tests. These

guidelines describe two types of test user qualifications: (a) generic qualifications that serve as a

basis for most of the typical uses of tests and (b) specific qualifications for the optimal use of tests

in particular settings or for specific purposes. They are *aspirational* because they identify

qualifications for the optimal use of tests in a competent and responsible manner. These

guidelines describe qualifications that apply to a variety of testing settings and for multiple

purposes; therefore, it is unlikely that a single test user possesses all the qualifications described

here. The qualifications should also be considered in relation to the context, setting, and purpose

of test use.

The guidelines apply most directly to standardized tests, such as measures of ability, aptitude, achievement, attitudes, interests, personality, cognitive functioning, and mental health. These guidelines apply to psychological tests whether or not they are administered by paper-and-pencil or electronically and whether or not they are scored and interpreted by a test user or electronically. The guidelines do not apply to unstandardized questionnaires and unstructured behavior samples or to teacher- or trainer-made tests to evaluate performance in education or training.

Various activities included in the testing process may be appropriately conducted by different people working collaboratively. Each participant should possess the knowledge, skills, and abilities relevant to his or her role. For example, different individuals may be responsible for deciding what constructs, conditions, or characteristics need to be assessed; selecting the appropriate tests; administering and scoring tests; and interpreting and communicating the results. Moreover, some testing activities may involve tasks that require limited professional knowledge (e.g., administering or scoring some tests, communicating simple test results). In such circumstances, test use should be <u>directed</u> by a qualified test user. It is this test user to whom these guidelines apply.

Persons whose psychological test use is confined to research will find that the degree to which these guidelines apply to their work depends on their research focus and the research setting. The sections that address knowledge and skills in relation to psychometrics, statistics, test administration, and scoring are applicable to research that uses psychological tests. When research is conducted with clinical populations or in settings where there are likely to be real or perceived implications for the test taker, additional guidelines may be applicable.

Testing and assessment. The use of psychological tests should typically be viewed within the context of the broader concept of assessment. Psychological assessment is a complex activity requiring the interplay of knowledge of psychometric concepts with expertise in an area of professional practice or application. Assessment is a conceptual, problem-solving process of gathering dependable, relevant information about an individual, group, or institution in order to make informed decisions. This process of data gathering and decision making involves a number of activities, including the following:

1. Recognizing the nature of the decisions to be made or the questions to be addressed;

2. Deciding what information is needed to answer these questions;

3. Selecting appropriate methods for acquiring this information, including tests, interviews, observations, surveys, or other data-gathering techniques;

4. Competently administering and scoring the selected tests according to standardized procedures when available and appropriate;

5. Accurately interpreting information, which may include knowing when to question the usual interpretation of a procedure because of intervening or mitigating circumstances;

6. Using assessment data and resultant interpretation to make a professionally sound decision; and

7. When appropriate, communicating assessment results in a way that is understandable to the client.

Many problems or questions to be addressed through assessment must be approached with a recognition of the potential for multiple coexisting or competing explanations. Such recognition

comes from the professional knowledge and judgment associated with advanced professional training and experience and not just from the ability to administer and score a particular test or other assessment instruments. This decision-making process is best conducted or directed by a professional with expertise in psychological assessment in a particular testing context.

Historical Background

In delineating the knowledge and skills important for the use of tests, the Task Force reviewed recent U.S. and international efforts to develop guidelines on test user qualifications. Several national and international professional organizations whose members use tests have addressed the issue of test user qualifications over the years. Their efforts are described briefly below.

Efforts in the United States. The APA appears to have been one of the first groups concerned with test user qualifications. The APA formed the Committee on Ethical Standards for Psychology in the late 1940s to develop its first set of ethical principles. The first topic in these ethical standards addressed the sale and distribution of psychological tests and diagnostic aids (Hobbs, 1951). The Committee released the ethical standards for the distribution of psychological tests in 1950. The complete set of ethical standards was adopted in 1953 (Golann, 1970). Since 1950, the APA has addressed the issue of test user qualifications broadly in subsequent revisions of its ethical principles (APA, 1981, 1992). The current version of the APA's ethical principles (APA, 1992) contains a number of standards that are related to appropriate test use, including specific principles related to the boundaries of competence for psychologists and the appropriate application and use of psychological assessment techniques.

Other professional groups that use psychological tests also have promulgated ethical

guidelines (e.g., American Association for Counseling and Development [now the American

Counseling Association; ACA], 1988; American Association for Marriage and Family Therapy,

1998; National Association of School Psychologists [NASP], 1992; National Council on

Measurement in Education [NCME], 1995). Indeed, the ACA has a specific set of

Responsibilities of Users of Standardized Tests (American Association for Counseling and

Development, 1988), popularly known as the RUST document. This document suggests that the

qualifications of test users depend on four factors: (a) the role of the user (e.g., administration

and scoring), (b) the setting, (c) the nature of the test, and (d) the purpose of testing.

In addition to developing its own ethical principles on test use by psychologists, the APA

has participated in formulating standards on the development and use of psychological and

educational tests (APA, American Educational Research Association [AERA], & NCME, 1954,

1966, 1974; AERA, APA, & NCME, 1985, 1999). The 1954 *Technical Recommendations for*

Psychological Tests and Diagnostic Techniques and the 1966 *Standards for Educational and*

Psychological Tests and Manual both referred to a categorization of test user qualifications first

approved by APA's Council of Representatives in 1950. The policy was referred to as the

"Ethical Standards for the Distribution of Psychological Tests and Diagnostic Aids" (APA,

1950) and included a three-level system for classifying test user qualifications.

This three-tiered system labeled some tests *Level A* (e.g., vocational proficiency tests)

and designated them as appropriate for administration and interpretation by nonpsychologists.

The next level of tests (e.g., general intelligence tests and interest inventories) was

labeled *Level B*. Qualifications for administering them included "some technical knowledge of

test construction and use, and of supporting psychological and educational subjects such as

statistics, individual differences, the psychology of adjustment, personnel psychology, and

guidance" (APA, 1950, p. 622). Over time, however, all those sanctioned "by an established

school, government agency, or business enterprise" (APA, 1950, p. 622) were reclassified as

eligible test users of Level B tests. Subsequent evidence suggests that those institutions did not

provide the oversight necessary to ensure that these test users were in fact qualified (Eyde et al.,

1993).

Finally, qualifications for the use of Level C tests (e.g., individually administered tests of

intelligence, personality tests, and projective methods) restricted their use to "persons with at

least a Master's degree in psychology, who have had at least one year of supervised experience

under a psychologist" (APA, 1950, p. 622). The Level C qualification also had some exceptions.

The reference to the three-tiered system was dropped from the 1974 (and subsequent)

Standards without a replacement, but casual inspection of test publishers' current catalogs

reveals that it is still in widespread use (cf. Robertson & Eyde, 1986).

An attempt to define test use was undertaken by an interdisciplinary group beginning in

1985. In that year, the APA, the AERA, the NCME, and test publishers formed the Joint

Committee on Testing Practices (JCTP).[1] The TUQWoG, a subgroup of the JCTP, immediately

set about developing a data-based approach to promoting good test use. TUQWoG conducted

several empirical studies designed to elucidate the types of competence problems exhibited by

---

[1]Before the work of the Test User Qualifications Working Group (TUQWoG) was
completed, the ACA and the American Speech-Language-Hearing Association had joined the
JCTP. The NASP has subsequently joined. The original requirement that one of the two
representatives of each organization be employed by a test publisher was dropped.

test users. The results of these studies were reported in a 1988 publication entitled *Test User*

Qualifications: A Data-Based Approach to Promoting Good Test Use (Eyde, Moreland,

Robertson, Primoff, & Most, 1988); in an article by Moreland, Eyde, Robertson, Primoff, and

Most (1995); and in a book of case studies (Eyde et al., 1993).

Despite all these efforts, evidence suggests that most of the problems associated with test

use are related to the competence of individual test users, although the uneven quality of test

construction and the ease with which test instruments can be obtained from some test publishers

also contribute to these problems (Tyler, 1986). In devising the present set of guidelines, the

TFTUQ kept in mind the types of problems identified by the empirical research and the

conclusion that much of the difficulty lies with test users. Thus, these guidelines were formulated

primarily to address characteristics of test users. This document *does not* pertain to the

development of tests, which is addressed in the *Standards for Educational and Psychological*

Testing (AERA, APA, & NCME, 1999).

International efforts. The Task Force found that concern over the misuse of tests has been

growing in the international psychology community over the past few years. Several countries

and international groups, including the International Test Commission (ITC), the British

Psychological Society (BPS), and the Canadian Psychological Association (CPA), have launched

initiatives to address concerns about test user qualifications.

In 1992, the CPA released a report on the adequacy of typical safeguards used by test

publishers to limit test access to qualified individuals (Simner, 1994). The report suggested that

test publishers did not uniformly apply the system of classifying tests according to three levels.

Some publishers did not use the three-tier system to screen test users, and those who did often

did not agree on the qualifications required for a particular test. In fact, there was disagreement

on the classification of about two thirds of the tests (Simner, 1994). The CPA report contained

recommendations for improving safeguards to protect the public from test misuse. These

recommendations ranged from replacing or supplementing the test-rating system used by the

publishers to requiring all first-time test users to complete a qualifications statement.

The BPS implemented a competence-based approach to certify test users (BPS, 1995,

1996). To date, the BPS certification system has focused on testing in occupational settings,

although the system may ultimately be expanded to address test user qualifications in educational

and health care settings as well. In the BPS system, test users are evaluated by assessors,

overseen by BPS-appointed verifiers, for demonstrated competence to use tests appropriately.

Those who are judged competent can apply for the BPS certificate in test competence and are

listed in a register that can be used by those purchasing testing services.

Finally, the Council of the ITC adopted international guidelines on core standards for test

use at its June 1999 meeting in Graz, Austria (ITC, 2000). The aim of this ITC project included

the production of a set of competencies (i.e., knowledge, abilities, and skills) for test use by

psychologists and nonpsychologists who use tests. The ITC guidelines represent the work of

specialists in psychological and educational testing (i.e., psychologists, psychometricians, test

publishers, and test developers) from the United States, Canada, Australia, and Europe.

APA's Role in Defining Test User Qualifications

The reason that the APA has sought to develop and promulgate guidelines for the use of

psychological tests evolves from a number of sources. As described above, historically, the APA

has recognized the need for and devoted considerable attention to the development of test user

qualifications. Although professionals from a variety of disciplines develop and use tests,

graduates from doctoral programs in psychology and educational and psychological

measurement have provided significant contributions to the science of testing and assessment.

The discipline of psychology is the historical root for psychological testing and provides the

research evidence and professional training to advance competent psychological assessment. The

APA formed the TFTUQ in 1996 in the belief that previous efforts, although useful, did not

provide the kind of specific guidance that many APA members and others were seeking. It is

appropriate for the discipline of psychology to establish guidelines for the proper use of

psychological tests.

II. Core Knowledge and Skills for Test Users

This section addresses the knowledge and skills that are important when test users make

decisions or formulate policies that directly affect the lives of test takers. The knowledge and

skills listed in this section are generic; however, the level of skill and depth of knowledge in

these areas may vary depending on the testing purpose and context. The next section describes

additional knowledge and skills that are relevant to the purpose or context in which tests are

used. These generic qualifications, in combination with the context-relevant qualifications

described later, are important for optimal test use. The *Standards for Educational and*

Psychological Testing (AERA, APA, & NCME, 1999) is an excellent resource for more

information on many of the concepts presented below.

1. Psychometric and Measurement Knowledge

It is important for test users to understand Classical Test Theory and, when

appropriate or necessary, Item Response Theory (IRT). The essential elements of

Classical Test Theory are outlined below. When test users are making

assessments on the basis of IRT, such as adaptive testing, they should be familiar

with the concepts of Item Parameters (e.g., item difficulty, item discrimination,

and guessing), Item and Test Information Functions, and Ability Parameters (e.g.,

theta).

1.1   *Descriptive statistics.* Test users should be able to define, apply, and interpret

concepts of descriptive statistics. For example, means and standard deviations are

often used in comparing different groups on test scales, whereas correlations are

frequently used for examining the degree of convergence and divergence between

two or more scales. Similarly, test users should understand how frequency distributions describe the varying levels of a behavior across a group of persons.

Test users should have sufficient knowledge and understanding of descriptive statistics to select and use appropriate test instruments, as well as score and interpret results. The most common descriptive statistics relevant to test use include the following:

1.1.1   Frequency distributions (e.g., cumulative frequency distributions)

1.1.2   Descriptive statistics characterizing the normal curve (e.g., kurtosis, skewness)

1.1.3   Measures of central tendency (e.g., mean, median, and mode)

1.1.4   Measures of variation (e.g., variance and standard deviation)

1.1.5   Indices of relationship (e.g., correlation coefficient)

1.2   *Scales, scores, and transformations.* Test results frequently represent information about individuals' characteristics, skills, abilities, and attitudes in numeric form. Test users should understand issues related to scaling, types of scores, and methods of score transformation. For example, test users should understand and know when to apply the various methods for representing test information (e.g., raw scores, standard scores, and percentiles). Relevant concepts include the following:

1.2.1   Types of scales

a.   Nominal scales

b.   Ordinal scales

       c.   Interval scales

       d.   Ratio scales

1.2.2   Types of scores

       a.   Raw scores

       b.   Transformed scores

            i.   Percentile scores

            ii.   Standard scores

            iii.   Normalized scores

1.2.3   Scale score equating

1.2.4   Cut scores

1.3   *Reliability and measurement error.* Test users should understand issues of test score reliability and measurement error as they apply to the specific test being used, as well as other factors that may be influencing test results. Test users should also understand the appropriate interpretation and application of different measures of reliability (e.g., internal consistency, test–retest reliability, interrater reliability, and parallel forms reliability). Similarly, test users should understand the standard error of measurement, which presents a numerical estimate of the range of scores consistent with the individual's level of performance. It is important that test users have knowledge of the following:

1.3.1 Sources of variability or measurement error

       a.   Characteristics of test taker (e.g., motivation)

b.   Characteristics of test (e.g., domain sampling, test length, and test

heterogeneity)

c.   Characteristics of construct and intended use of test scores (e.g.,

stability of characteristic)

d.   Characteristics and behavior of test administrator (e.g., importance

of standardized verbal instructions,)

e.   Characteristics of the testing environment

f.   Test administration procedures

g.   Scoring accuracy

1.3.2 Types of reliability and their appropriateness for different types of tests and

test use

a.   Test–retest reliability

b.   Parallel or alternative forms reliability

c.   Internal consistency

d.   Scorer and interrater reliability

1.3.3 Change scores (or difference scores)

1.3.4 Standard error of measurement (i.e., standard error of a score)

1.4   *Validity and meaning of test scores.* The interpretations and uses of test scores,

and not the test itself, are evaluated for validity. Responsibility for validation

belongs both to the test developer, who provides evidence in support of test use

for a particular purpose, and to the test user, who ultimately evaluates that

evidence, other available data, and information gathered during the testing process

to support interpretations of test scores. Test users have a larger role in evaluating

validity evidence when the test is used for purposes different from those

investigated by the test developer.

Contemporary discussions of validity have focused on evidence that

supports the test as a measure of a construct (sometimes called *construct validity*).

For example, evidence for the uses and interpretations of test scores may come

through evaluation of the test content (content representativeness), through

evidence of predictions of relevant outcomes (criterion-related validity), or from a

number of other sources of evidence. Test users should understand the

implications associated with the different sources of evidence that contribute to

construct validity, as well as the limits of any one source of validity evidence.

1.4.1   Types of evidence contributing to construct validity

      a.   Content

      b.   Criterion related

      c.   Convergent

      d.   Discriminant

1.4.2   *Normative interpretation of test scores.* Norms describe the distribution of

test scores in a sample from a particular population. Test users should

understand how differences between the test taker and the particular

normative group affect the interpretation of test scores.

      a.   Types of norms and relevance for interpreting test taker score (e.g.,

            standard scores and percentile norms)

  b.   Characteristics of the normative group and the generalizability

          limitations of the normative group

  c.   Type of score referent

          i.   Norm referenced

          ii.   Domain referenced (criterion referenced)

          iii.   Self-referenced (ipsative scales)

  d.   Expectancy tables

2. Selection of Appropriate Test(s)

To select the best test or test version for a specific purpose, test users should have

knowledge of testing practice in the context area and the most appropriate norms when more

than one normative set is available. Knowledge of test characteristics such as psychometric

properties (presented above), basis in theory and research, and normative data (where

appropriate) should influence test selection. For example, normative data or decision rules may

not be accurate when (a) important characteristics of the examinee are not represented in the

norm group, (b) administration or scoring procedures do not follow those used in standardizing

the test, (c) characteristics of the test may affect its utility for the situation (e.g., ceiling and floor

effects), (d) the test contains tasks that are not culturally relevant to the test taker, or (e) the

validity evidence does not support decisions made on the basis of the test scores.

Test users should have an understanding of how the construction, administration, scoring,

and interpretation of tests under consideration match the current needs. Mismatches in these

dimensions between the selected test and the current testing situation represent important factors

that should be considered and which may invalidate usual test interpretation.

More specifically, to select an appropriate test for a particular use, it is important that test users understand and consider the following:

2.1   Intended use of the test score

2.2   Knowledge of the method and procedures used to develop or revise the test being considered

    2.2.1   Definition of the construct that the test purports to measure

    2.2.2   Definition of the test purpose and its intended context of use

    2.2.3   Type of keying or scaling used

        a.   Rational or theoretical

        b.   Empirical

        c.   Internal consistency or construct homogeneity (e.g., factor analysis)

    2.2.4 Scoring procedures (e.g., clinical, mechanical, and correction for guessing)

    2.2.5 Type of score interpretation

        a.   Criterion or domain referenced

        b.   Norm referenced

        c.   Ipsative

    2.2.6 Item and scale score characteristics

        a.   Item format

        b.   Difficulty level

        c.   Reliability (e.g., internal consistency and test–retest)

2.2.7 Validity evidence of test scores

    a. Construct validity evidence

        i.   Content representativeness

        ii.   Criterion related

        iii.   Convergent

        iv.   Discriminant validity

        v.   Cross-validation

        vi.   Validity generalization (e.g., the effects of sample size, test and criterion reliability and range restriction , and dichotomization of variables)

        vii.   Criterion Characteristics (e.g., sufficiency, relevance)

2.2.8 Test bias (see 4.2 below for details)

2.2.9 Description of validation, normative, and/or standardization group(s)

    a.   Characteristics of groups (such as age, gender, race, culture, language, disabilities, geographic region, socioeconomic status [SES], educational or grade level, motivational set, mental status, and item format familiarity)

    b.   Sample size(s)

    c.   Recency of data

2.2.10 Test administration procedures

    a.   Standardization procedures

    b.   Time limits (power vs. speed)

2.3   Knowledge of test taker variables that may moderate validity and interpretation of

scores (such as age, gender, race, culture, language, disabilities, geographic

region, era or time period tests, SES, educational or grade level, motivational set,

mental status, and item format familiarity)

2.4   Other or special requirements and limitations of test

2.5   Adequacy of the match between test characteristics and present need in terms of

the following:

2.5.1   Construct measured

2.5.2   Difficulty level

2.5.3   Validity

2.5.4   Reliability

2.5.5   Test bias

2.5.6   Normative data

2.5.7   Similarity of normative group with present group

2.5.8   Test administration procedures

a.   Accommodations for disabilities (when appropriate)

b.   Characteristics of test administrator

c.   Adaptation for individuals with different primary language (when

appropriate)

2.5.9   Special requirements and limitations of test

3. Test Administration Procedures

Knowledge about procedural requirements, confidentiality of test information, communication of results, and test security is important for many testing applications, as is familiarity with standardized administration and scoring procedures and understanding a test user's ethical and legal responsibilities and the legal rights of test takers. Similarly, it is important that test users understand the legal and ethical issues related to the release of test materials, including issues of confidentiality, depending on the context of the testing and the characteristics of the test taker.

Test users should be able to explain test results and test limitations to diverse audiences. Written communications should include the purpose of the test and the setting in which the testing occurred. In preparing written reports on test results, test users should be aware that the test scores might become separated from the interpretive report over time.

More specifically, test users should be familiar with the following:

3.1   Legal rights of test takers

3.2   Standardized administration procedures

3.3   Scoring procedures

3.4   Confidentiality of test materials and test information

   3.4.1   Safeguards for protecting test materials

      a.   Protection against copyright infringement

      b.   Protection against unauthorized dissemination of test

           items/keys/scoring procedures

       3.4.2   Safeguards for protecting protocols and test results

            a.   Legal issues

            b.   Ethical issues

   3.5   Reporting results to the test taker, caregiver, or others as appropriate

       3.5.1   Characteristics of meaningful reports

       3.5.2   Amount of information to report

       3.5.3   Legal and ethical issues

4. Ethnic, Racial, Cultural, Gender, Age, and Linguistic Variables

Consideration of these variables may be important to the proper selection and use of psychological tests. For certain purposes, legal requirements influence or restrict the testing, scoring, interpretation, analysis, and use of test data of individuals in different subgroups. In some cases (e.g., employment testing), the use of gender, race, and/or ethnicity in test interpretation is illegal. Test users should consider and, where appropriate, obtain legal advice on legal and regulatory requirements to use test information in a manner consistent with legal and regulatory standards. Issues associated with testing individuals from particular subgroups, such as race or ethnicity, culture, language, gender, age, or other classifications, are addressed in greater detail in the 1999 version of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999).

The APA's promulgated *Guidelines and Principles for Accreditation of Programs in Professional Psychology* (APA, 1996) discussed the need for psychology training programs to address issues of cultural diversity. The APA demonstrated its interest in and sensitivity to these issues by establishing the Commission on Ethnic Minority Recruitment and Training in

Psychology. In addition, the Task Force on Delivery of Services to Ethnic Minority Groups, under the auspices of the Board of Ethnic Minority Affairs, published *Guidelines for Providers of Psychological Services to Ethnic, Linguistic, and Culturally Diverse Populations* (APA, 1990). These guidelines were approved by the APA's Council of Representatives. In addition, the ITC has issued "Guidelines for Adapting Educational and Psychological Tests: A Progress Report" (Hambleton, 1994), which provides recommendations about adapting tests for cross-cultural testing.

For test users using tests with different ethnic, racial, cultural, gender, and language groups, knowledge of the following is important:

4.1 *Construct equivalence.* Test users strive to be familiar with the literature regarding issues of construct equivalence (e.g., cultural equivalence) in its various forms and how this might affect the selection, use, and interpretation of psychological tests for individuals whose dominant language is not the language of the test or who are from different racial, ethnic, or gender groups.

    4.1.1   Information concerning the influence of psychological characteristics (e.g., motivation, attitudes, and stereotype threat) on test performance

    4.1.2   Orientations and values that may alter the definition of the constructs(s) being assessed and how those factors may affect the interpretation of test results

    4.1.3   Requirements of the testing environment and how that may affect the performance of different groups

4.2   *Test bias.* Test users should be familiar with the legal and psychometric literature

pertaining to test bias for different racial, ethnic, cultural, gender, and linguistic

groups and how this might affect decisions pertaining to selection of tests and

interpretation of test results. It is important that test users know the following:

4.2.1   Laws and public policies concerning use of tests that may have

implications for test selection, as well as administration and interpretation

4.2.2   Procedures for examining between-groups differences in test performance

4.2.3   Empirical literature concerning differential validity for racial or cultural

groups

5. Testing Individuals with Disabilities

Tests are administered to increasing numbers of persons with disabilities in a variety of

settings and for a multitude of purposes. The requirement to accommodate an individual with a

disability in the testing situation raises many complex issues for test users. Test users must

frequently make decisions regarding the use of tests that were not developed and normed for

individuals with disabilities. In such circumstances, confidence in the inferences drawn from test

results may be diminished. There may be legal requirements concerning the accommodation of

individuals with disabilities in test administration and the use of modified tests. Test users should

consider and, where appropriate, obtain legal advice on legal and regulatory requirements

regarding appropriate administration of tests and use of test data.

Several efforts were initiated during the mid-1990s to provide guidance to test users for

assessing individuals with disabilities. The APA Task Force on Test Interpretation and Diversity

published a book identifying the scientific and policy issues related to the interpretation of tests

used with individuals for whom the tests were not developed, standardized, and validated

(Sandoval, Frisby, Geisinger, Scheuneman, & Grenier, 1998). This text identified important

considerations in assessing individuals with specific types of disabilities (e.g., deafness,

blindness, and learning disabilities). Additionally, a working group of the JCTP published a

sourcebook on assessing individuals with disabilities. This sourcebook for practitioners describes

some of the pertinent legal and regulatory information, as well as types of accommodations,

required documentation, and the use of tests in various contexts (e.g., employment, admissions,

and counseling; Ekstrom & Smith, in press). Finally, the 1999 *Standards for Educational and*

Psychological Testing (AERA, APA, & NCME, 1999) includes a chapter on technical

considerations for testing individuals with disabilities. Those who administer tests to individuals

with disabilities should be familiar with the legal, technical, and professional issues governing

the use of tests with individuals with disabilities, including the following:

> 5.1 *Legal issues.* Test users involved in assessing individuals with disabilities should be
>
> familiar with the relevant legal requirements and enforcement guidance for assessing
>
> individuals with disabilities for specific purposes (e.g., Section 504 of the
>
> Rehabilitation Act, the Individuals With Disabilities Education Act, and the
>
> Americans With Disabilities Act) and obtain legal advice in these matters where
>
> appropriate.
>
> 5.2 *Test selection.* Test users should possess the knowledge to make an appropriate
>
> selection of measures. Test users strive to have current information regarding
>
> availability of modified forms of the tests in question.

5.3 *Test accommodation.* Test users strive to be familiar with the available literature addressing the various accommodations appropriate for individuals with disabilities and, to the extent available, on the effects of test accommodation on test score interpretation and use. When there is a need to modify a test, test users should have the knowledge and skills needed to modify the test appropriately for the test taker while maintaining all feasible standardized features and to communicate those modifications as appropriate.

5.4 *Interpretation of test results of individuals with disabilities.* Test users strive to be familiar with the literature regarding how external factors and characteristics associated with the disability may affect the interpretation of test scores, such as the following:

    5.4.1   Effects of the testing environment and the tests being used on the performance of individuals with disabilities

    5.4.2   Inferences based on the test scores accurately reflect the construct, rather than construct-irrelevant, characteristics associated with the disability

    5.4.3   Knowledge of whether regular norms or special norms are appropriate for the characteristic in question

6. Supervised Experience

In addition to having certain knowledge and skills needed for appropriate test use, it is important that test users have the opportunity to develop and practice their skills under the supervision of appropriately experienced professionals. This supervision typically begins in graduate school and continues throughout training until any credentials that are necessary to

practice independently have been attained. The structure and focus of supervision will vary depending on the domain(s) in which supervision is being administered. Because testing is conducted by psychologists with different specialties, as well as by nonpsychologists, this report cannot prescribe a specific format or mechanism for supervision. However, focused and setting-specific supervision of sufficient intensity and duration is important for those engaged in testing.

7. Summary of Core Knowledge and Skills

The intent of this section is to delineate the multiple domains and competencies important for users of psychological tests. Although qualifications may vary by practice area, a combination of high-level skill and professional judgment is important. The test user's key function is to make valid interpretations of test scores and data, often collected from multiple sources, using proper test selection, administration, and scoring procedures. To provide valid interpretations, it is important that test users be able to integrate knowledge of applicable psychometric and methodological principles, the theory behind the measured construct and related empirical literature, the characteristics of the particular tests used, and the relationship between the selected test and the particular testing purpose, the testing process, and, in some contexts, the individual test taker.

III. Test User Qualifications in Specific Contexts

The context in which psychological tests are used includes both the setting and the purpose of testing. Test user qualifications vary across settings, as well as within settings, depending on the purpose of testing. This section addresses the context-relevant qualifications that build on the generic qualifications described above.

Regardless of the setting, psychological tests are typically used for the following purposes:

1. *Classification*. To analyze or describe test results or conclusions in relation to a specific taxonomic system and other relevant variables to arrive at a classification or diagnosis.

2. *Description*. To analyze or interpret test results to understand the strengths and weaknesses of an individual or group. This information is integrated with theoretical models and empirical data to improve inferences.

3. *Prediction*. To relate or interpret test results with regard to outcome data to predict future behavior of the individual or group of individuals.

4. *Intervention planning*. To use test results to determine the appropriateness of different interventions and their relative efficacy within the target population.

5. *Tracking*. To use test results to monitor psychological characteristics over time.

This section describes five major contexts in which tests are commonly used: employment, educational, vocational/career counseling, health care, and forensic. There also may be other contexts that require specific qualifications. The qualifications important in the major contexts, as well as appropriate training and supervision, are discussed below.

Employment Context

Many employers use tests as part of the assessment process to develop work-related information and recommendations or decisions about people who work for them or are seeking employment with them. Test users in this context should have not only the qualifications identified as core knowledge and skills but also an understanding of the work setting, the work itself, and the worker characteristics required of the work situation. They strive to know what skills, abilities, or other individual difference characteristics enable people to perform effectively (as defined in a variety of ways) in a particular work setting. Test users consider the strengths and weaknesses of different methods for determining the human requirements of the work situation and how to conduct such job, work, or practice analyses. They also should consider and, where appropriate, obtain legal advice about employment law and relevant court decisions (see Dunnette & Hough, 1990, 1991, 1992, 1994; Guion, 1998).

Some persons who administer tests and communicate test results in an employment setting (e.g. Human Resources personnel and recruiters) may not be considered test users by this document. This document applies to those who select tests for use and determine how test results are to be used in employment decision-making. Under this scenario, the test user may be a company employee, a test vendor employee, or a consultant.

Classification. Organizations seek to classify or place people in jobs to maximize overall utility to both the individuals and the institution. To perform these activities well, test users strive to be knowledgeable about job clustering (e.g., creation of job families), validity, cost-benefit analysis, utility analysis, and measurement of work outcomes (Alley, 1994; Bobko, 1994; Zeidner & Johnson, 1994).

Psychological tests are sometimes used to certify people as qualified to perform certain job or work activities. Test takers unable to pass these certification tests are deemed unqualified at present to perform particular tasks, activities, or jobs at a defined level of competence and may not be eligible to practice the profession or perform those tasks. Test users should have knowledge of the task or work and knowledge of the level of performance required for competent practice. This means that test users define the task or criterion, measure the required knowledge and skills, and identify the required performance level (i.e., set cut scores that reflect the level of task, skill, and knowledge required for competent practice). They strive to have a thorough knowledge of job, work, or practice analysis and of content validation principles and strategies (Knapp & Knapp, 1995; K. Schmitt & Shimberg, 1996).

Description. Description of an individual's current abilities, skills, interests, personality, knowledge, or other personal characteristics can be a significant part of the assessment process in industrial, business, or governmental settings concerned with human resources management. This information is the starting point for determining the fit between an individual and work in a given setting; identifying areas of needed individual, team, or organizational development; providing feedback about likely success in different work activities and settings; planning career choices and paths; and auditing organizational or unit readiness. Those who use psychological tests to describe individual, team, or organizational characteristics in the employment setting should consider information about the work and its setting. Thus, knowledge about job, work, or career analysis is important (see Campion, 1994; Dawis & Lofquist, 1984; Fleishman & Quaintance, 1984; Gael, 1988; Goldstein, Zedeck, & Schneider, 1993; Hall, 1986; Peterson, Mumford, Borman, Jeanneret, & Fleishman, 1999).

Prediction. Psychological tests may be used as part of a larger assessment process to help make predictions about an individual's future training performance, job performance, trustworthiness, attrition, or a variety of other work-related criteria. These predictions are often made to facilitate recommendations or decisions about selection, promotion, or succession planning.

Test users involved in testing to predict future employment criteria make every effort to be knowledgeable about the work setting and the work itself and, hence, job or work analysis methods. They understand the principles of psychological measurement as they apply to tests and as they apply to criteria. They also should understand performance measurement, criterion constructs and their measurement, relationships between various predictor constructs and criterion constructs, research methods and design, validity concepts and evidence, test bias, adverse impact analysis, utility analysis, validity generalization, and group differences, and consider and, where appropriate, obtain legal advice regarding applicable collective bargaining and contract requirements, federal and state guidelines on employment testing, employment law, and relevant court decisions (see Anderson & Herriot, 1997; Campbell, 1996; Campbell & Campbell, 1988; Cascio, 1990, 1991; Dunnette & Hough, 1990, 1991; Guion, 1998; Hakel, 1998; Howard, 1995; Murphy, 1996; N. Schmitt & Borman, 1993). Those who use tests for selection, promotion, and succession planning purposes should also be aware of motivational set and its possible effect on applicant responses and the validity of inferences based on them (Anastasi, 1988; Hough, 1998).

Intervention planning. Employment testing may be part of an analysis of the test taker's training and development needs. Test results may provide information for developing plans to

improve skill and performance of current work responsibilities and anticipated work responsibilities. Test results may also be used as part of career planning activities. When tests are used for these purposes, test users make every effort to be knowledgeable about such matters as the work itself, the work setting, performance appraisal and performance measurement, criterion constructs and their measurement, training and development, career development, coaching and mentoring, and training needs analysis (Goldstein, 1989; Hall, 1986; London, 1995; Ostroff & Ford, 1989).

Employment testing may be part of an outplacement process. If testing is done as part of an involuntary process that determines who is to be retained and who is to be laid off, test users should be knowledgeable about the work itself and the work setting (hence, job, work, or practice analysis methods), performance measurement, criterion constructs and their measurement, validity concepts and evidence, test bias, adverse impact analysis, and group differences, and consider and, where appropriate, obtain legal advice regarding collective bargaining and contract requirements applicable to the particular organization or work setting, federal and state guidelines on employment testing, employment law, and relevant court decisions (see Arvey & Faley, 1988; Colarelli & Beehr, 1993; Guion, 1998; Kozlowski, Chao, Smith, & Hedlund, 1993; Landy & Farr, 1983; Murphy & Cleveland, 1995). If testing is done as part of an outplacement, voluntary job search process, test users should be knowledgeable about vocational and career guidance, job loss, and labor markets (see Caplan, Vinokur, Price, & van Ryn, 1989; Dawis, 1991; Dawis & Lofquist, 1984; Hall, 1986; Holland, 1976; Pickman, 1994).

Employment testing may also be a part of a monitoring system designed to identify individuals who are at risk for performing below an acceptable level. The individuals may be

employed in sensitive-duty (high cost for mistakes) jobs. Airline pilots, nuclear power plant operators, and undercover police officers or agents are examples. Those who use tests to identify at-risk individuals should have the qualifications listed under the *Classification* and *Prediction* sections above. When the assessment of risk involves the identification of psychopathology or other health issues, the test user qualifications described in the Health Care Context section below also apply.

Tracking. Psychological tests may be used in predictive, criterion-related validation studies in which individuals and their performance are tracked over time. In addition to the knowledge recommended for the use of psychological tests for prediction purposes (see the *Prediction* section above), test users who track individuals or their performance also need to understand how task or work performance and criterion performance requirements may change over time (Ackerman, 1987; Borman, 1991; Fleishman & Fruchter, 1960; Ghiselli, 1956; Kane, 1986; Komaki, Collins, & Penn, 1982). In addition, test users who conduct reassessments should be familiar with the effects of repeated use of assessment procedures on both the individual and the findings obtained. For example, frequent retesting of a skill might appear advisable but could produce practice effects and spuriously inflated results, unless alternative forms of the tests are available (Chall & Curtis, 1990).

Training and supervision. Training for test use in the employment context is best obtained by successful completion of an integrated program of study that includes industrial psychology, psychology of individual differences, measurement theory, job/work/practice analysis, performance measurement, and employment law relevant to the testing situation. Experience and supervision using tests in settings similar to those in which employment tests are

used is important. For test users who provide assessment of health outcomes or understanding of health problems of individuals and groups (e.g., those working in employee assistance programs [EAPs]), the qualifications described in the Health Care Context section below also apply.

Educational Context

The results of psychological tests often serve as relevant information to guide educational decisions about both students and programs (AERA, APA, & NCME, 1999). This section addresses the use of psychological tests to assess educational outcomes or educational processes pertaining to an individual, a group of individuals, or an educational institution. Psychological tests are used in a variety of educational settings, including preschools, elementary and secondary schools, colleges, universities, technical schools, business training programs, counseling centers, health and mental health settings that offer educational services, and educational consulting practices. Psychological tests are typically used to acquire information about students to make informed decisions about such issues as student admissions and placement, educational programming, student performance, and teacher or school effectiveness. Given the wide range of educational settings and the multiple uses for group and individual test data, it is likely that more individuals are administered tests in an educational context than in any other setting (Bersoff, 1979, 1999).

On an individual level, psychological tests are often used to describe a student's learning or behavioral strengths and weaknesses. The results may then be used to develop educational interventions, to determine appropriate educational placements (e.g., special education, gifted education, magnet school program, or alternative educational setting), or as part of clinical diagnostic assessment to guide therapeutic services (Fagan & Wise, 1995).

Assessment of groups of individuals, often called *large-scale testing*, typically addresses

questions or concerns about educational programs or policies (Hambleton & Jurgensen, 1990).

Decision makers may aggregate results from psychological tests and use this information to

evaluate program effectiveness and develop recommendations for changes to educational

programs or systems. Test users in these cases may use standardized tests or nonstandardized

procedures (e.g., performance events or portfolios of student work) to obtain information about

cognitive ability or academic achievement levels of a group of students (Fuchs & Fuchs, 1990).

A majority of states require students to complete large-scale test batteries to determine their

proficiency relative to state standards. In some instances, results from such large-scale tests are

reported only at the aggregate level, providing district, school, or classroom results. In other

instances, results are reported for individual students as well as districts, schools, or classrooms.

The qualifications described above in the section on Core Knowledge and Skills for Test

Users apply to individuals using psychological tests in an educational context. Topics that have

particular relevance in educational settings include the representativeness of the test sample,

attention to language and cultural diversity, and the use of cut scores in selection for special

programs (Henning-Stout & Brown-Cheatham, 1999; Kranzler, 1999; Reynolds & Kamphaus,

1990; Salvia & Ysseldyke, 1995). Test users should also understand the cognitive and emotional

factors that affect student learning, as well as the social and political factors that affect schools as

learning environments (Gettinger & Stoiber, 1999; Medway & Cafferty, 1999; Tharinger &

Lambert, 1999; Ysseldyke & Elliott, 1999). Those who use psychological tests in social

institutions like schools should be particularly skilled at communicating the results of testing to

many different audiences, including educational decision makers, teachers, students, parents, and the public (AERA, APA, & NCME, 1999; Illback, Zins, & Maher, 1999).

The specific nature of the qualifications that are important to test use in the educational context depend on both the purpose for which tests are used (e.g., classification or prediction) and the level of focus (e.g., individual or large-scale testing). The knowledge, skills, and abilities associated with optimal test use at both the individual and group level are described in relation to the purpose for which the test is used in an educational context.

Classification. Tests are often used to identify or classify individual students or groups of students for admission to special programs. In public elementary and secondary schools, the most frequently used formal classification system is probably the one used to determine eligibility for special education services as required by federal and state law (e.g., the Individuals With Disabilities Education Act). Therefore, test users in educational contexts should consider and, where appropriate, obtain legal advice regarding state and federal laws related to the provision of educational and related services to disabled students (Jacob-Timm & Hartshorne, 1998; Reschly & Bersoff, 1999). Many schools also use curriculum-tracking schemes (e.g., general vs. college preparatory classes), which categorize and then place students in separate instructional tracks or ability groupings, each with its own eligibility criteria. Schools also use classification systems to identify individuals at risk for school failure, eligible for gifted and talented programs, or for admission to magnet programs. Individuals using psychological tests for classification purposes, both in individual and large-scale assessments, should be familiar with the taxonomic systems used by schools and other educational settings as well as the

psychometric limitations of the tests used (Kamphaus, Reynolds, & Imperato-McCammon, 1999;

Macmann & Barnett, 1999; Reynolds, 1990).

Test users should also possess the knowledge to select instruments that are appropriate

for the characteristics of the student being evaluated (AERA, APA, & NCME, 1999). For

example, tests that have adequate reliability and validity for assessing school-age students may

be inappropriate for use with preschool children (Bracken, 1987, 1994; Nuttall, Romero, &

Kalesnik, 1999). If a test has been developed, normed, and validated for use with individuals

from one language, culture, race, or ethnic group, it may not be appropriate for individuals from

other cultural or ethnic populations (Figueroa, 1990). For individual assessment, test users

consider and, when appropriate, integrate information from multiple sources, such as

psychological and educational test data, behavioral observations and ratings, school records, and

interviews with parents and teachers (Salvia & Ysseldyke, 1995).

Large-scale tests are used for a variety of purposes, including program accountability and

decisions related to admissions and educational placement. In most instances, important

decisions about students should not be based on a student's performance on a single test (AERA,

APA, & NCME, 1999). When schools, districts, or states develop or select a test to determine

student achievement relative to state standards, test users should have the skills and knowledge

to determine the degree of correspondence among the standards, curricula, and test content.

When critical decisions, such as graduation or retention, are based on test results, test users strive

to consider students' opportunity to learn the stated content and identify other sources of relevant

data that reflect student proficiency. When tests are used for college placement, test users

determine the degree of alignment between the test's content and the college courses, as well as

understand the relationship between predicted and actual performance in subsequent courses before determining a cut score or other classification criteria. Legal requirements may influence or restrict the use of rank ordering or cut scores, particularly if these practices have a disproportionate effect on one or more subgroups.

Description. Psychological tests are also used in educational settings to describe aspects of learners' skills and abilities, such as learning styles, motivation, reading readiness, and emotional maturity. These student characteristics may be assessed to describe a student's academic strengths or weaknesses or to differentiate educational approaches based on individual need. Group measures of interests, attitudes, cognitive abilities, or emotional adjustment may also provide a basis for interventions designed to remediate current problems or to prevent future difficulties.

Large-scale assessments are often used by schools, districts, and states to measure the general level of student performance. Often such test use is designed to evaluate the effects of curricular decisions or program outcomes. In some instances, schools or teachers may be held accountable for their students' test results, with penalties imposed for scores below expectations. Therefore, it is important that test users attend to the multiple factors that contribute to test score differences between schools, classrooms, or districts (e.g., student motivation, quality of prior educational experiences, and parental support of educational goals).

Prediction. In the educational context, tests are often used to predict the future behavior or academic success of a student or group of students. In individual assessment, tests are often used to screen students for placement in special programs (e.g., gifted education, programs for students at risk of educational or behavioral problems, and magnet programs for special interests

or abilities) or to place them in an instructional group or track based on a prediction of expected future performance.

In large-scale testing, admissions tests are required for entry into most undergraduate, graduate, and professional programs. These tests help the institution estimate the students' readiness for an academic program and provide a means to compare the academic preparation of students who have attended different schools, who have completed different courses, and who have been graded according to different criteria. Admissions tests are also useful in college counseling, providing students with useful information on their potential for academic success at different colleges and universities. In addition, most colleges use specially developed placement tests to determine a student's eligibility for particular courses.

Whether the focus of the assessment is an individual student or a group of students, the test user should recognize that each student's future performance is affected by many factors. In addition to examining a student's abilities, characteristics, and motivation, test users should have the skills and knowledge to evaluate the relative contribution of teacher competence and motivation, school and classroom climate, peer group influence, class size, and other factors that play a critical role in determining a student's future performance (Gettinger & Stoiber, 1999). They also strive to understand the likely course of learning difficulties and developmental variations in the acquisition of academic skills (Tharinger & Lambert, 1999). Finally, test users strive to be familiar with the literature on how group differences (e.g., ethnicity, gender, race, and SES) may affect performance on standardized tests, grades, school completion, and other outcomes that may be used in predicting academic success (Figueroa, 1990; Henning-Stout & Brown-Cheatam, 1999).

Intervention planning. Psychological tests are frequently used to plan interventions for one student or a group of students. Psychological tests are commonly used as part of the individual diagnostic assessment of students with learning or behavioral problems (Kamphaus et al., 1999; Salvia & Ysseldyke, 1995; Woody, La Voie, & Epps, 1992). The results from these tests help to describe or diagnose the educational strengths and weaknesses of students or their behavioral difficulties and contribute to the development of educational, behavioral, or mental health interventions. Test users involved in intervention planning for individual students strive to be knowledgeable about alternative instructional approaches; school curriculum; special education services; and therapeutic interventions, such as counseling, group dynamics, and behavioral interventions (Hughes, 1999; Shapiro & Cole, 1994). Those who use tests to prescribe interventions based on assessed student characteristics should be familiar with the empirical evidence for using test data to make such decisions.

Test results sometimes provide a rationale for educational interventions that affect a large number of students, such as a modification in instructional approach (Algozzine & Ysseldyke, 1992; Gettinger & Stoiber, 1999; Illback et al., 1999). One example is the decision to replace a phonics approach in reading instruction with a whole-language approach. Test users strive to clearly communicate to decision makers the appropriateness of inferences based on test data and the likely effects of program changes on various groups of students. Test results may also be used as a basis for individual interventions, such as removing a student from school (e.g., school suspension) or placing that individual in a private residential program for severely disturbed or impaired individuals. Here, test users should consider how significantly a change in educational placement may affect a student's self-concept, educational achievement, and overall well-being

(Jacob-Timm & Hartshorne, 1998; Woody et al., 1992). Test users should consider and, where appropriate, obtain legal advice about relevant state and federal laws dealing with changes in placement and the use of educational interventions that affect school placement (Jacob-Timm & Hartshorne, 1998; Reschly & Bersoff, 1999) as well as the legal protections afforded to parents and students, including, where applicable, due process rights and requirements of informed consent (Jacob-Timm & Hartshorne, 1998).

Tracking. Test users in school settings often administer tests multiple times to track the effects of educational programming or interventions. In individual assessment, special education law requires that students classified as disabled be reassessed at least every 3 years so that students are given a periodic review of their status (Jacob-Timm & Hartshorne, 1998; Reschly & Bersoff, 1999; Salvia & Ysseldyke, 1995). Even students who are not classified as having a disability but who receive a modification in their educational programming are reassessed periodically to determine if the interventions are having the desired outcomes and are still warranted.

Groups of students may be assessed yearly to document academic progress or to evaluate a program's effectiveness (Algozzine & Ysseldyke, 1992; Illback et al., 1999). Such aggregated student data are frequently used as the basis for modifying instructional programs and policies. In some cases, the school is required to obtain evidence of program effectiveness to receive continued funding for that program.

When tests are used for tracking purposes in educational settings, test users should understand the effects of repeated test administrations on the students and on the findings obtained. For example, frequent retesting of reading achievement to guide instruction might

appear advisable but could produce serious practice effects and spuriously inflated results, unless

alternative forms of the reading tests are available (Chall & Curtis, 1990; Shapiro & Elliott,

1999). Those who use tests to track student performance also strive to be aware of the social and

instructional context variables that may influence student performance, so that changes in test

scores are not automatically attributed to changes in student abilities (Greenwood, Carta, &

Atwater, 1991; Ysseldyke & Elliott, 1999).

Training and supervision. In addition to the knowledge, skills, and abilities outlined for

all test users, the user of psychological tests in the educational context should be knowledgeable

in the content areas of educational and psychological theory and practice, as well as the legal

requirements and protections for test takers that are relevant to the type of assessment being

conducted. This combination of generic psychometric knowledge and context-relevant expertise

is best acquired in an integrated program of advanced professional preparation, such as that

acquired in a doctoral program in school or educational psychology or educational measurement.

As noted earlier, the type of training and the breadth and depth of knowledge in each of these

domains may vary for different test users depending on whether they are responsible for

individual diagnostic testing or large-scale testing. Test users in an educational environment

should possess an appropriate practice credential where such credential is legally required to

provide the type of testing being offered. It is also important that they receive supervised

experience appropriate to their role and setting in the use of tests to address educational problems

or questions.

Individuals using psychological tests to place children in special education programs

should be knowledgeable in areas such as developmental and social psychology, diagnostic

decision making, child psychopathology, and special education practices. They should consider and, where appropriate, obtain legal advice on special education law. Furthermore, test users involved in individual diagnostic testing strive to be competent in communicating and translating assessment results into educationally relevant and appropriate recommendations that are likely to result in meaningful improvement.

Individuals using psychological tests to address large-scale testing questions related to admissions, student grouping, or instructional programming should be particularly knowledgeable in the domains dealing with psychometrics, instructional design, educational and developmental psychology, and measurement theory. In addition, individuals doing large-scale testing or research in school settings should be knowledgeable and skilled in communicating the results of tests to diverse audiences including school personnel, students, parents, policymakers, the media, and the public in general. Individuals using tests for college or graduate school admissions, for counseling, or for placement also strive to be knowledgeable about the empirical evidence related to using tests to make such decisions in higher education. They should consider and, where appropriate, obtain legal advice regarding the legal protections for test takers in higher education settings.

Career/Vocational Counseling Context

Psychological testing in the career/vocational counseling context is used to help people make appropriate educational, occupational, retirement, and recreational choices and to assess difficulties that impede the career decision-making process. Career/vocational counselors integrate their knowledge of career demands with information about beliefs, attitudes, values, personalities, mental health, and abilities, with the goal of promoting beneficial career

development, life planning, and decision making. Successful career adjustment is based on occupational, intellectual, personal, developmental, educational, and societal factors. Information about values, interests, abilities, achievements, mental health, and work experience is important to the process (Zunker, 1990). The individual's self-knowledge about values, strengths, weaknesses, motivation, psychological characteristics, and interests are also relevant (Herr & Cramer, 1996).

Career/vocational testing overlaps somewhat with employment testing, but the two often serve different purposes. In employment testing, typically the job is already defined, whereas in career/vocational psychology tests are used to help individuals make personally relevant career choices. Another distinction is that in employment testing the client is the employer (not the test taker), whereas in career/vocational testing the client is usually the test taker, even when a parent or school is financially responsible for the testing. Another distinction between the fields is that there are many more legal issues governing the use of psychological tests in personnel selection than there are in career/vocational assessment.

Psychological tests in the career/vocational counseling context are used to help individuals make decisions about career and life planning. Testing can provide persons knowledge about their work-related and avocational interests, their abilities, and their values and help them understand how these fit into the existing opportunities and requirements of the workplace and into their leisure activities. Along with the knowledge, skills, and abilities identified earlier, test users strive to understand how individuals' particular interests, values, abilities, and skills relate to their choice of work and leisure activities. Test users also should have substantive knowledge in related areas of psychology, such as adolescent and adult

development, personality, and psychopathology, as well as detailed and current knowledge of measurement questions involved with assessing interests, abilities, personality dimensions, and values. Test users should be able to integrate complex results that cross these multiple domains of assessment.

Test users also make every effort to be knowledgeable about types of work settings, work cultures and values, and the characteristics and requirements of types of jobs. They strive to integrate the results of multiple measures from a number of different domains with their knowledge of vocational theories (Osipow & Fitzgerald, 1996) and career taxonomies (Holland, 1997; Lowman, 1991).

Test users identify and work with individual difference and systemic variables that may influence the person–environment fit. Such factors include the individual's family system, gender, ethnicity, cultural background, physical ability, SES, and psychological problems. Test users should be able to recognize and work not only with the problems explicitly presented by the test taker but also with other problems, including underlying emotional difficulties or environmental impediments that could affect the way the test taker uses test results. For example, a test taker's family or cultural background might deem certain careers unacceptable and therefore require the test user to process this perception and assist in generating viable vocational options.

Often the person seeking career or leisure counseling is experiencing a life transition that brings additional personal, developmental, and emotional stress. In addition, such individuals may struggle with emotional problems that make deciding on a career difficult. For example, those who lack self-esteem and confidence may find it challenging to engage in self-assessment,

reflect on the world of work, and confidently select an occupation. To deal effectively with such complex mixtures of career, developmental, and emotional concerns, vocational test users should have qualifications similar to those required in the health care context (see the Health Care Context section below).

Classification. The primary focus of vocational classification is to identify an individual's career-related skills, abilities, and characteristics (e.g., interests and personality factors) and then match them with the requirements of specific jobs or job categories. Vocational classification may also be used to match an individual with a specific school or program or to help a person identify satisfying leisure activities or outlets for prized abilities.

Knowledge of individual differences in cognition and personality are central in the assessment of person–environment fit. Career/vocational counselors may administer cognitive, achievement, and aptitude tests to determine a test taker's skills or special competencies (Kapes, Mastie, & Whitfield, 1994; Lowman, 1991). Differential patterns of abilities may be as important as scores on individual ability measures, so testing may need to cover a wide range of competencies. Career/vocational counselors may use personality inventories, interest inventories, and other assessment procedures to help them understand the test taker's preferences, values, learning history, and occupational or leisure goals. By effectively communicating test results to test takers, career/vocational counselors help their clients to better understand the fit of their characteristics with their environment.

Description. Similar to the health care context, a holistic description of the individual's personality and mental health is important in the career/vocational counseling context (Gysbers, Heppner, & Johnston, 1999). The coexistence and interaction of career and mental health

problems (Blustein & Spengler, 1995; L. Lucas & Epperson, 1988; M. S. Lucas, 1992) support

the need for test users to assess personality and mental health problems that may impede

successful career development. Moreover, test users may want to assess important constructs,

such as career indecision and career choice anxiety, with those who have a history of difficulty in

vocational decision making. Thus, test users in the career/vocational counseling context should

be qualified to assess the mental health functioning of individuals seeking career counseling in

order to determine the most effective approach (refer to the following section on Health Care

Context).

Prediction. Prediction is often a central concern for vocational assessors. That is, the

results of a variety of vocational tests are assumed to reflect stable, enduring traits that are

relevant to future work performance and satisfaction. Although related constructs such as

interests and cognitive abilities demonstrate stability over a period of years, the degree of

consistency partly depends on the developmental level of the test taker. For example, students

may lack the experience necessary to crystallize vocational interests until they have reached

college age (Blustein, Pauling, DeMania, & Faye, 1994; Tinsley & Barrett, 1977). Vocational

test users should temper predictions of future behavior with the knowledge that test takers'

further development and specific situations may strongly influence their work behaviors.

Intervention planning. In some cases, the vocational intervention consists entirely of the

administration and interpretation of tests and the communication of assessment findings. This is

often true when the test taker's increased self-knowledge regarding interests, values, personality,

and the world of work is the goal of the intervention. In these cases, test users strive to engage

the individual actively in the process of test interpretation (Tinsley & Bradley, 1986).

In other cases, additional vocational interventions may be used in conjunction with psychological assessment (Fouad, 1994) or may be identified as needed by the career assessment process. For example, testing may yield a list of potentially suitable occupations that the test taker can investigate and experience in internships and part-time work. Or the test results may indicate a lack of differentiation among the test taker's vocational interests, suggesting that additional experience is needed before more specific work preferences can be developed. To perform effective career/vocational interventions, test users should have knowledge of career development theories and skills in interviewing and history taking, as well as knowledge of relevant educational and career information resources. Test users strive to be aware of discriminatory patterns that exist in various careers.

In some cases, evaluation of test results shows that further psychological intervention is needed. Test users should be able to evaluate patterns of behavior and test results, recognize test takers who will not be able to benefit from vocational information because of significant developmental, cognitive, emotional, or physical problems, and treat or refer them appropriately.

Tracking. Tests used for career and vocational assessment may provide standards against which to compare patterns of subsequent growth or deterioration. Tests may be useful, on an individual level, in revealing patterns of change after traumatic or remediative experiences. Grouped test data can provide important information for uses such as determining the characteristics of employees in occupations or organizations or students in particular majors and how they may change over time. Test users should be knowledgeable about the psychometric and context-related implications of assessing career development over time.

Training and supervision. The use of psychological tests in career and vocational

assessment, as described above, requires complex skills in career and mental health assessment,

not just simply learning to use tests in isolation. Appropriate training (e.g., that obtained through

doctoral programs in relevant areas of psychology) includes coursework in adolescent and adult

development, as well as the domain of vocational/career psychology. Test users engaged in

career counseling and testing should be knowledgeable about measurement theory, as described

earlier. They strive to be skilled in involving clients in the interpretation of vocational tests.

Finally, it is important that their training include supervised experience in the use of

psychological tests in vocational/career settings, and relevant experience in educational,

counseling, health care, and occupational settings.

Health Care Context

Health care is the provision of services to individuals who seek help in enhancing their

physical or mental well-being or in dealing with behaviors, emotions, or issues that are

associated with suffering, disease, disablement, illness, risk of harm, or risk of loss of

independence. Health care assessment commonly occurs in private practice, rehabilitation,

medical or psychiatric inpatient or outpatient settings, schools, EAPs, and other settings that

address health care needs.

Psychological tests are used as part of the assessment process to develop health-related

information and recommendations or decisions about people to improve their physical or mental

health. Those who use tests for this purpose should have thorough grounding both in the core

knowledge and skills enumerated earlier and in the specialized knowledge, training, or

experience of specific substantive areas of health care. With so many specialized areas in health

care, it is impractical to specify any single set of core knowledge requirements, technical

competencies, or supervised training experiences for test users. More detailed guidance is often

provided in the guidelines and standards developed by professionals working in a specialized

health care area.

Because health care providers' decisions and actions can have very important and

sometimes very dramatic effects on the lives of the people they serve, the health care profession

is heavily regulated. Test users should keep abreast of ethical standards relative to psychological

assessment (Bersoff, 1999; Koocher, 1993; Koocher & Keith-Spiegel, 1998) as well as

regulations and laws at both state and federal levels on such subjects as confidentiality, duty to

warn, mandated reporting, and patient rights (APA Committee on Legal Issues, 1996; APA

Committee on Psychological Tests and Assessment, 1996; Koocher, Norcross, & Hill, 1998) and

obtain legal advice in these matters where appropriate.

In the health care context, psychological test data are typically used to augment

information gathered from other sources (e.g., patient and collateral interviews, behavioral

observations, and laboratory results). Health care providers who use psychological tests strive to

effectively integrate results from multiple tests and sources of information. Psychological test

users strive to understand how the nature of the setting (e.g., psychiatric hospital) and the

characteristics of test takers (e.g., those who have a physical illness or disability or who are on

medication) might affect the process of test administration, the results, and the interpretation.

Test users strive to communicate the technical aspects of their findings to other professionals as

well as to health care consumers in language that is appropriate and understandable to each.

Classification. When psychological tests are used for classification purposes, the most

common goal is the assignment of a mental health, medical, or other diagnosis. In these

instances, psychological test findings are generally combined with interview and historical data,

behavioral observations, and data from other sources to derive a formal diagnosis. When

diagnosis is the goal of testing, test users combine the skills associated with competent testing

with a separate set of knowledge, skills, and experiences related to classification and diagnosis in

the population of interest.

Test users should be able to identify and evaluate factors that may influence diagnostic

determinations and that are frequently not accounted for in the development, standardization, and

norming of psychological tests. For example, when working with persons whose physical

symptoms may affect test performance, test users should be knowledgeable about and

experienced at distinguishing illness-related test results from other determinants for a person's

test performance (e.g., motivation, demographics, personality traits, or other medical

considerations).

Test users seek to understand determinants of diagnostic accuracy in relation to both the

specific assessment procedures being used and the decisions that need to be made. For example,

when psychological tests are used to screen for specific health problems such as alcoholism or

dementia, test users should consider how fluctuations in base rates in different populations may

affect the sensitivity and specificity of test results (Ivnik et al., in press).

Description. Psychological tests are also used in health care to provide a more

comprehensive description of individuals by delineating their unique personality, emotional,

cognitive, or other characteristics. For example, a combination of personality, academic,

aptitude, interest, and cognitive tests may be used to help describe the areas of both preserved and compromised functioning for a young person who is in a rehabilitation facility in hope of returning to work after suffering a head injury in a motor vehicle accident (MVA). When performing primarily descriptive assessments in health care, test users should consider the construct validity of the tests that they select and how these constructs are manifested in day-to-day behavior. To avoid misinterpreting normal inter- and intratest variance as pathology, test users who work in health care should consider the limits of normal variance when different psychological characteristics are simultaneously measured. When individuals are followed over time and psychological tests are repeated one or more times, test users are attentive to issues that relate to how "meaningful change" is distinguished from normal test–retest variability (Ivnik et al., 1999; Jacobson & Truax, 1991; Sawrie, Chelune, Naugle, & Luders, 1996).

Prediction. Health care professionals are frequently asked to make predictions (i.e., prognoses) about the persons they serve, and psychological test users may specifically be asked to make testing-based predictions. For example, the person who tested the MVA victim mentioned above may be asked to "predict" when this person might return to work or to school or the person's final level of recovery. In these instances, test users strive to be knowledgeable about the predictive limits of testing. When tests are used to make predictions in health care settings, test users strive to understand the patient's unique characteristics (e.g., personality features, special strengths, disabilities or disorders, and sociocultural issues), the natural course of medical conditions, the likely efficacy of planned interventions, and relevant base-rate information. Test users strive to understand the empirical evidence of a test's ability to make accurate predictions. For example, neuropsychologists who make predictions about a person's

need for assistance in daily activities should know how well their test instruments predict

relevant functional capacities (Lemsky, Smith, Malec, & Ivnik, 1996).

Intervention planning. In health care settings, data from psychological tests may be used

in planning interventions. Intervention planning refers to the selection of specific remediation

activities based on a thorough knowledge of the problem being addressed and available treatment

options. Test users involved in intervention planning may use tests to provide information on an

individual's particular problem (classification), strengths and weaknesses (description), and the

efficacy of treatment options (prediction). The same set of knowledge and skills required for

competent classification, description, and prediction are also important in the development of an

optimal treatment plan. For example, personality tests may be used to modify treatment

approaches in a therapeutic setting (Maruish, 1999). Because intervention planning involves a

specific type of prediction (i.e., the likelihood that a patient will benefit from a particular form of

treatment), test users strive to be aware of the limitations discussed above related to prediction

and the scientific evidence supporting available treatments.

Tracking. In some circumstances, multiple sequential administrations of the same test(s)

are frequently needed to document how psychological characteristics change over time or as a

consequence of treatment (e.g., to track the course of a patient's illness or recovery). To interpret

these results, test users strive to be knowledgeable about how repeated exposures to test

procedures and test content influence subsequent test performances (e.g., practice effects),

including how conditions (e.g., memory deficits) present during one examination may affect the

results of later testing. Test users also strive to understand how to distinguish measurement error

from reliable test score change (e.g., Ivnik et al., 1999; Jacobson & Truax, 1991; Sawrie et al.,

1996). Psychological tests are sometimes used to measure treatment outcome. For example, test results may help determine eligibility for health care services or to monitor treatment efficacy. If this application is different from the test's original purpose, test users should be aware of potential factors that may limit the usefulness or validity of the test data as an indicator of treatment outcome.

Training, supervision, and licensure. In the health care context, the qualifications described above are best obtained through doctoral training in psychology, which includes psychological testing supervision in one or more health care settings that are similar to the setting(s) in which a specific test user intends to practice. The APA's model licensing act (APA, 1987) recommends for health care psychologists that state credentialing bodies require 2 years of full-time supervised experience with a minimum of 1 hr/week of individual supervision provided by an appropriately credentialed professional. Also, guidelines for training programs such as the APA's *Guidelines and Principles for the Accreditation of Programs in Professional Psychology* (APA, 1996) include requirements for supervised experience in graduate training, predoctoral internship, and mandated postdoctoral supervision. Finally, some health care specialties have defined the core knowledge, training, and supervised experiences that are needed for fully competent test use (e.g., neuropsychology; Hannay et al., 1998). The specific health care setting in which a test user works (e.g., mental health facilities or EAPs) will define the added content areas that a test user should master.

In addition to coursework in psychological testing, personality theory and assessment, and measurement theory, independent health services providers who use tests for health care needs should be particularly knowledgeable in psychopathology, health psychology, life-span

developmental psychology, and the biological bases of behavior. Test users in the health care context should also be skillful in clinical diagnostic interviewing and familiar with mental health diagnostic and classification systems. As noted earlier, the breadth and depth of knowledge in each of these domains, as well as additional technical qualifications, may vary depending on the specific area of specialized functioning.

Health care professionals who use psychological tests are credentialed by the state or province in which they work. Credential renewal in many states requires documentation of continuing professional education. Those who use psychological tests in a health care context strive to obtain knowledge, supervised training, and professional experiences that go beyond the profession-specific knowledge, training, and experiences they obtained during graduate education, practica, internship, residency, or fellowship. For some test users whose original graduate education and training were not in clinical areas, graduate-level respecialization programs can provide additional education and training.

Forensic Context

In forensic settings, psychological tests are used to gather information and develop recommendations about people who are involved in legal proceedings. Test users in forensic settings should possess a working knowledge of the functioning of the administrative, correctional, or court system in which they practice. They strive to be familiar with the statutory, administrative, or case law in the specific legal context where the testing occurs or, where appropriate, obtain legal advise on the pertinent laws. They strive to communicate test results in a way that is useful for the finder of fact (i.e., the judge, the administrative body, or the jury). This includes communicating verbally with lawyers, writing formal reports, and giving sworn

testimony in deposition or court.

The problems encountered in forensic settings are varied and complex, often involving medical illnesses, developmental problems, and multiple forms of psychopathology, so test use often requires expertise in multiple health care areas. This section addresses those who use clinical, rehabilitation, and neuropsychological tests in legal contexts, as well as those who believe that their test data will serve as a foundation for legal consultation or testimony. Thus, in addition to the core qualifications identified earlier, the qualifications described above for test users in health care contexts typically apply to test users in forensic settings.

This section does not address test use by two groups of experts who also may work in forensic settings. Specifically, this section does not apply to those who use psychological tests to conduct research in applied areas of forensics, such as memory, social psychology, or human factors. Nor does it apply to those who use tests in applied areas, such as clinical, rehabilitation, or neuropsychological practice or industrial/organizational or educational psychology and who may be asked to provide consultation or testimony based on their training, education, or experience about work with their clients. However, these test users should be sensitive to the potential ramifications of assuming multiple roles (Greenberg & Shuman, 1997; Shuman, Greenberg, Heilbrun, & Foote, 1999).

Those who use tests for forensic purposes should possess substantive knowledge in areas of psychology related to the forensic issues. For example, in correctional or criminal settings, knowledge about violence, criminality, and the relationship of psychopathology to those behaviors and activities is germane (Heilbrun et al., 1998). Similarly, when assessing families in child custody or parental rights cases, it is important for test users to understand family

dynamics, parenting, and different forms of child custody (APA Committee on Professional

Practice and Standards, 1994).

Assessments for forensic purposes often occur in outpatient, inpatient, and correctional

settings (Heilbrun, 1992; Melton, Petrila, Poythress, & Slobogin, 1997). Each of these settings

exerts specific influences that may significantly alter how tests are administered and interpreted.

For example, in correctional settings test users strive to understand how the test results may be

affected by the level of privacy of the testing location, the noise in the area, and even the degree

of objective danger and threat to the inmate from other residents. Further, test users strive to be

knowledgeable about the effect of incarceration on the presentation of psychopathology, possible

effects of the trial or litigation process on client presentation, and the assessment of response set

issues (Rogers, 1997).

Classification. Diagnostic skills are important for the use of psychological tests in forensic

settings. In most situations, the assessment will include multiple measures to provide a thorough

and legally defensible diagnosis (Heilbrun, 1992; Heinze & Grisso, 1996). Thus, test users in

forensic settings strive to integrate results from multiple tests with knowledge of accepted

diagnostic taxonomies (e.g., the *Diagnostic and Statistical Manual of Mental Disorders* [4th ed.;

American Psychiatric Association, 1994) and knowledge about how test findings relate to these

systems (Talge, 1995).

Test users strive to identify and evaluate critical factors that may influence diagnostic

determinations. Among these factors are the defendant's response set and the effects of

incarceration and litigation on the defendant's test results. A thorough knowledge of response set

and its influence on test results (Rogers, 1997) may be needed for accurate interpretation of test

results.

Because of the high stakes in legal proceedings (monetary settlements, child custody, jail

sentences, and even the death penalty), test takers may be motivated to exaggerate or minimize

their symptoms. Because diagnosis may be complicated by these response biases, test users in

forensic settings strive to recognize these factors and account for them in the interpretation.

Additionally, test users in forensic settings should understand that psychopathology as measured

by tests may be improved or exacerbated by incarceration and that trial proceedings and litigation

may affect test data by increasing or decreasing the litigant's anxiety, depression, or anger

(Weissman, 1991).

Test users are often required to evaluate historical information to help the court arrive at a

determination of causation or to review events that have occurred in the past to ascertain whether

those events relate in some way to a legal standard. Consultation with family members or friends

of the examinee may also add to the accuracy of the interpretation of test results.

For example, in criminal settings, test users may be asked to assist the court in

determining whether the defendant was criminally responsible for his or her behavior at the time

of the offense. Or a test user may be asked to assess the defendant's capacity to waive his or her

Fourth and Fifth Amendment (Miranda) rights—critical for determining whether a confession is

admissible in court (Grisso, 1986). In tort (civil lawsuit) settings, determination of causation (the

legal nexus between a specific event and a psychopathological condition) is often a critical

element for determining whether even the minimum basis for a lawsuit exists. Even in contexts

where causation involves strictly technical knowledge from other fields (e.g., chemistry or

physiology), test users may be asked to provide legally admissible information on the

psychological or neuropsychological status of an examinee without attributing causation.

Those using tests in forensic settings to determine the causation of legally relevant conditions or events strive to be knowledgeable about how the tests are used to determine the origins or natural histories of mental disorders. Users of neuropsychological tests may use patterns of scores on those tests to inform opinions about the cause of specific behaviors (e.g., Martzke, Swan, & Varney, 1991; Varney & Menefee, 1993). Assessment of brain trauma or toxic chemical reactions may fall into this category. Test users assessing traumatic emotional reactions should have knowledge about the relationship of specific score patterns with specific types of emotional trauma. Test users also should have knowledge of relevant epidemiological studies (Kilpatrick & Resnick, 1993; Swanson, 1994) and etiology of mental conditions.

Description. In forensic settings, clients are described in relation to a legal standard in a particular context. The most obvious example is the application of the standards for legal competency (to stand trial, to execute a legal document, and to be executed). These standards are established by legislation and case law (see Grisso, 1986).

Standards are applied to clients for a variety of forensic purposes. In criminal cases, a major focus has been the assessment of individuals for determining criminal responsibility or insanity at the time of the offense (Rogers, 1986). In correctional settings, assessment results in conjunction with historical or behavioral data may determine whether an inmate is described as a high-, medium-, or low-security risk (see Megargee, 1979, 1994). In tort or disability settings, the standard may be a legal description of an emotional condition, which will be applied to examinees to determine their eligibility for compensation under administrative regulations (e.g., Social Security) or laws (Sales & Perrin, 1993).

To perform these descriptive activities, test users should consider and, where appropriate, obtain legal advice on the applicable legal standard to craft the appropriate assessment strategy to produce a legally useful result and to interpret the assessment results in light of that standard (Heilbrun, 1992). Test users may be called upon to explain how the test data are relevant to the applicable legal standards.

Prediction. In forensic practice, test users are often asked to make a statement about the future behavior of a test taker (Otto, 1992). In civil commitment settings, for example, most states' criteria for involuntary commitment include the examinee's dangerousness to self or others (Monahan & Steadman, 1996). In criminal settings, statements concerning the examinee's potential for recidivism on parole from prison may be a critical element of a prerelease evaluation (Borum & Grisso, 1995; Webster, Douglas, Eaves, & Hart, 1997). In tort settings, predictions about the prognosis of an emotional condition may be necessary for determining damages in a lawsuit (Sales & Perrin, 1993). In domestic relations settings, predictions of a child's reaction to a specific custody arrangement may be a critical part of the custody evaluation.

To use test results for prediction, test user should be knowledgeable about the base rates of legally relevant behaviors (e.g., violence, suicide, or posttraumatic states) and the contribution of situational factors (e.g., life stresses, substance abuse, or treatment with psychotherapy or medication) to these behaviors.

Intervention planning. Intervention planning based on test data may be an important part of the test user's responsibilities in forensic settings. For example, in divorce, adoption, or abuse and neglect cases, recommendations for treatment for a child or family may be integral to the

child custody recommendation. In a sentencing evaluation, recommendations for treatment may be included in deliberations and influence the duration or location of the convicted person's incarceration. In tort settings, treatment recommendations may, in part, determine the amount of monetary compensation provided for the plaintiff.

In addition to the prediction skills indicated above, skills important for intervention planning in forensic settings include both knowledge of how test data may be helpful for selecting appropriate treatment strategies and knowledge of how test data may assist in predicting response to treatment.

Tracking. In forensic settings, it is often important to know how test data may be affected by the passage of time and by events that occur between repeated test administrations. In working with children, for example, test users should consider the effects of developmental sequences in the assessment of the child's current emotional condition to trace the origins of that condition to specific events such as traumatic experiences or changes in custody. Tests may assist in the process of ruling out alternative causes of conditions. Although the determination of causation is generally a classification activity (see the *Classification* section above), a test user may be called upon to review a sequence of test data generated through a series of testing periods. This is most likely to occur in cases where the test user has an opportunity to review test data that were gathered before the commission of a criminal offense or before the injury that is the focus of subsequent litigation. Such data may assist the test user in assessing issues of legal causation.

Training and supervision. The knowledge, skills, and abilities identified in this section are best obtained through doctoral training in psychology and relevant supervised experience, as

described in the Health Care Context section. Licensure requirements for those who use

psychological tests in the forensic context are similar to those required of practitioners in the

health care context.

The coursework and training for individuals who use tests in the forensic context are very

comparable with the coursework and training for those who use tests for health care needs,

although a basic introduction to psychology and the law is also desirable. In addition, training in

the specific area of law (e.g., criminal responsibility) may be important. This may be acquired

through formal or continuing education course work (Bersoff et al., 1997; Ogloff, Tomkins, &

Bersoff, 1996) or through mentoring by, or consultation with, someone trained and

knowledgeable in the relevant statutes (e.g., a lawyer specializing in the field in question).

Supervised experience in the conduct of a particular type of forensic evaluation may also be

critical. Experience in one forensic area (e.g., child custody evaluation) does not necessarily

prepare the test user for functioning in another forensic area (e.g., death penalty phase testimony;

Haas, 1993; White, 1987).

IV. A Look Forward

The psychological testing process has undergone significant technological change over the past few decades. The use of computers to administer tests and to score and interpret test results is already an important part of everyday testing. Emerging technologies of the Internet and other innovations that expand applications across vast distances may significantly alter the relationship of the test user, test taker, and the consumer of testing results.

Some of the positive changes resulting from these new technologies include wider availability, greater accuracy, and increased accessibility of tests. Continuing improvements in the development of interpretive algorithms and expert systems are leading to diminishing concurrent human oversight of the testing process. This technology will simplify some aspects of the assessment process. As the application of new technology to the testing process produces improved but more complex testing services, it may become necessary for the knowledge and skills articulated in this document to be supplemented with increased technological sophistication. Ironically, this increased complexity may mandate more extensive education and training in the fundamentals of test use. The knowledge and skills articulated here will become even more important as test users are required to distinguish technology-based style from science-based substance.

References

Ackerman, P. L. (1987). Individual differences in skilled learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102,* 3–27.

Aiken, L. S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate training in statistics, methodology and measurement in psychology: A survey of PhD programs in North America. *American Psychologist, 45,* 721–734.

Algozzine, B., & Ysseldyke, J. (1992). *Strategies and tactics for effective instruction.* Longmont, CO: Sopris West.

Alley, W. E. (1994). Recent advances in classification theory and practice. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 431–442). Hillsdale, NJ: Erlbaum.

American Association for Counseling and Development. (1988). *Responsibilities of users of standardized tests.* Washington, DC: Author.

American Association for Marriage and Family Therapy. (1998). *AAMFT code of ethics.* Washington, DC: Author.

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychological Association. (1950). Ethical standards for the distribution of psychological tests and diagnostic aids. *American Psychologist, 5,* 620–626.

American Psychological Association. (1981). Ethical principles of psychologists. *American Psychologist, 36,* 633–638.

American Psychological Association. (1987). Model act for state licensure of psychologists. *American Psychologist, 42,* 696–703.

American Psychological Association. (1990). *Guidelines for providers of psychological services to ethnic, linguistic, and culturally diverse populations*. Washington, DC: Author.

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist, 47,* 1597–1611.

American Psychological Association. (1996). *Guidelines and principles for accreditation of programs in professional psychology*. Washington, DC: Author.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1954). *Technical recommendations for psychological tests and diagnostic techniques*. Washington, DC: Author.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1966). *Standards for educational and psychological tests and manual*. Washington, DC: Author.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1974). *Standards for educational and psychological tests*. Washington, DC: Author.

American Psychological Association Committee on Legal Issues. (1996). Strategies for private practitioners coping with subpoenas or compelled testimony for client records of test data. *Professional Psychology: Research and Practice, 27,* 245–251.

American Psychological Association Committee on Professional Practice and Standards. (1994). Guidelines for child custody evaluations in divorce proceedings. *American Psychologist, 49,* 677–680.

American Psychological Association Committee on Psychological Tests and Assessment. (1996). Statement on disclosure of test data. *American Psychologist, 51,* 644–668.

Anastasi, A. (1988). *Psychological testing* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Anderson, N., & Herriot, P. (Eds.). (1997). *International handbook of selection and assessment.* West Sussex, England: Wiley.

Arvey, R. D., & Faley, R. H. (1988). *Fairness in selecting employees* (2nd ed.). Reading, MA: Addison-Wesley.

Bersoff, D. N. (1979). Regarding psychologists testily: Legal regulation of psychological assessment in public schools. *Maryland Law Review, 39,* 27–120.

Bersoff, D. N. (1999). *Ethical conflicts in psychology* (2nd ed.). Washington, DC: American Psychological Association.

Bersoff, D. N., Goodman-Delahunty, J., Grisso, J. T., Hans, V. P., Poythress, N. G. Jr., &

Roesch, R. G. (1997). Training in law and psychology: Models from the Villanova conference.

American Psychologist, 52, 1301–1310.

Blustein, D. L., Pauling, M. L., DeMania, M. E., & Faye, M. (1994). Relation between

exploratory and choice factors and decision progress. *Journal of Vocational*

Behavior, 44, 75–90.

Blustein, D. L., & Spengler, P. M. (1995). Personal adjustment: Career counseling and

psychotherapy. In W. B. Walsh & S. H. Osipow (Eds.), *Handbook of vocational psychology:*

Theory, research, and practice (2nd ed., pp. 295–329). Mahwah, NJ: Erlbaum.

Bobko, P. (1994). Issues in operational selection and classification systems: Comments

and commonalities. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection*

and classification (pp. 443–456). Hillsdale, NJ: Erlbaum.

Borman, W. C. (1991). Job behavior, performance, and effectiveness. In M. D. Dunnette

& L. M. Hough (Eds.), *Handbook of industrial organizational psychology* (Vol. 2, pp. 271–326).

Palo Alto, CA: Consulting Psychologists Press.

Borum, R., & Grisso, T. (1995). Psychological test use in criminal forensic examinations.

Professional Psychology: Research and Practice, 26, 465–473.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal

levels of technical adequacy. *Journal of Psychoeducational Assessment, 4,* 313–326.

Bracken, B. A. (1994). Advocating for effective preschool assessment practices: A

comment on Bagnato and Neisworth. *School Psychology Quarterly, 9,* 103–108.

British Psychological Society. (1995). *Certificate statement register: Competencies in occupational testing. General information pack (Level A).* (Available from the British Psychological Society, 48 Princess Road East, Leicester, England LEI 7DR)

British Psychological Society. (1996). *Certificate statement register: Competencies in occupational testing. General information pack (Level B).* (Available from the British Psychological Society, 48 Princess Road East, Leicester, England LEI 7DR)

Campbell, J. P. (1996). Group differences and personnel decisions: Validity, fairness, and affirmative action. *Journal of Vocational Behavior, 49,* 122–158.

Campbell, J. P., & Campbell, R. J. (Eds.). (1988). *Productivity in organizations: New perspectives from industrial and organizational psychology.* San Francisco: Jossey-Bass.

Campion, M. A. (1994). Job analysis for the future. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 1–12). Hillsdale, NJ: Erlbaum.

Caplan, R. D., Vinokur, A. D., Price, R. H., & van Ryn, M. (1989). Job seeking, re-employment, and mental health: A randomized field experiment in coping with job loss. *Journal of Applied Psychology, 74,* 759–769.

Cascio, W. F. (1990). *Applied psychology in personnel management* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Cascio, W. F. (1991). *Costing human resources: The financial impact of behavior in organizations* (3rd ed.). Boston: PWS-Kent.

Chall, J. S., & Curtis, M. E. (1990). Diagnostic assessment of reading. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (pp. 535–551). New York: Guilford Press.

Colarelli, S. M., & Beehr, T. A. (1993). Selection out: Firings, layoffs, and retirement. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 341–384). San Francisco: Jossey-Bass.

Dawis, R. V. (1991). Vocational interests, values, and preferences. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial organizational psychology* (Vol. 2, pp. 833–872). Palo Alto, CA: Consulting Psychologists Press.

Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment: An individual-differences model and its applications*. Minneapolis: University of Minnesota Press.

Dunnette, M. D., & Hough, L. M. (Eds.). (1990). *Handbook of industrial and organizational psychology* (Vol. 1). New York: Rand McNally.

Dunnette, M. D., & Hough, L. M. (Eds.). (1991). *Handbook of industrial and organizational psychology* (Vol. 2). New York: Rand McNally.

Dunnette, M. D., & Hough, L. M. (Eds.). (1992). *Handbook of industrial and organizational psychology* (Vol. 3). New York: Rand McNally.

Dunnette, M. D., & Hough, L. M. (Eds.). (1994). *Handbook of industrial and organizational psychology* (Vol. 4). New York: Rand McNally.

Ekstrom, R. B., & Smith, D. K. (Eds.). (in press). *Assessing individuals with disabilities: A guide for practitioners.* Washington, DC: American Psychological Association.

Eyde, L. E., Moreland, K. L., Robertson, G. J., Primoff, E. S., & Most, R. B. (1988). *Test user qualifications: A data-based approached to promoting good test use. Issues in scientific psychology* (Report of the Test User Qualifications Working Group of the Joint Committee on Testing Practices). Washington, DC: American Psychological Association.

Eyde, L. E., Robertson, G. J., Krug, S. E., Moreland, K. L, Robertson, A. G., Shewan, C. M., Harrison, P. L., Porch, B. E., Hammer, A. L., & Primoff, E. S. (1993). *Responsible test use: Case studies for assessing human behavior*. Washington, DC: American Psychological Association.

Fagan, T., & Wise, P. S. (1995). *School psychology: Past, present and future.* Columbus, OH: Charles E. Merrill.

Figueroa, R. A. (1990). Assessment of linguistic minority group children. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (pp. 671–696). New York: Guilford Press.

Fleishman, E. A., & Fruchter, B. (1960). Factor structure and predictability of successive stages of learning Morse code. *Journal of Applied Psychology, 44,* 97–101.

Fleishman, E. A., & Quaintance, M. K. (1984). *Taxonomies of human performance: The description of human tasks*. Orlando, FL: Academic Press.

Fouad, N. A. (1994). Annual review 1991–1993: Vocational choice, decision-making, assessment, and intervention. *Journal of Vocational Behavior, 45,* 125–176.

Fuchs, L. S., & Fuchs, D. (1990). Curriculum-based assessment. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (pp. 435–455). New York: Guilford Press.

Gael, S. (Ed.). (1988). *The job analysis handbook for business, industry, and government* (Vols. 1–2). New York: Wiley.

Gettinger, M., & Stoiber, K. C. (1999). Excellence in teaching: Review of instructional

and environmental variables. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school*

psychology (3rd ed., pp. 933–958). New York: Wiley.

Ghiselli, E. E. (1956). Dimensional problems of criteria. *Journal of Applied Psychology,*

40, 1–4.

Golann, S. E. (1970). Ethical standards for psychology: Development and revision, 1938–

1968. *Annals of the New York Academy of Sciences, 169,* 398–405.

Goldstein, I. L. (Ed.). (1989). *Training and development in organizations*. San Francisco:

Jossey-Bass.

Goldstein, I. L., Zedeck, S., & Schneider, B. (1993). An exploration of the job analysis–

content validity process. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in*

organizations (pp. 3–34). San Francisco: Jossey-Bass.

Greenberg, S. A., & Shuman, D. W. (1997). Irreconcilable conflict between therapeutic

and forensic roles. *Professional Psychology: Research and Practice, 28,* 50–57.

Greenwood, C. R., Carta, J. J., & Atwater, J. (1991). Ecobehavioral analysis in the

classroom: Review and implications. *Journal of Behavioral Education, 1,* 59–77.

Grisso, T. (1986). *Evaluating competencies: Forensic assessment and instruments*. New

York: Plenum.

Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*.

Hillsdale, NJ: Erlbaum.

Gysbers, N. C., Heppner, M. J., & Johnston, J. A. (1999). *Career counseling: Process,*

issues, and techniques. Boston: Allyn & Bacon.

Haas, L. J. (1993). Competence and quality of performance of forensic psychologists. *Ethics and Behavior, 3,* 251–266.

Hakel, M. D. (Ed.). (1998). *Beyond multiple choice: Evaluating alternatives to traditional testing for selection*. Mahwah, NJ: Erlbaum.

Hall, D. T. (Ed.). (1986). *Career development in organizations*. San Francisco: Jossey-Bass.

Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment, 10,* 229–244.

Hambleton, R. K., & Jurgensen, C. (1990). Criterion-referenced assessment of school achievement. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (pp. 456–476). New York: Guilford Press.

Hannay, H. J., Bieliauskas, L., Crosson, B. A., Hammeke, T. A., Hamsher, K. deS., & Koffler, S. (1998). Proceedings of the Houston Conference on Specialty Education and Training in Clinical Neuropsychology. *Archives of Clinical Neuropsychology, 13,* 157–249.

Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior, 16,* 257–272.

Heilbrun, K., Hart, S. D., Hare, R. D., Gustafson, D., Nunez, C., & White, A. J. (1998). Inpatient and postdischarge aggression in mentally disordered offenders: The role of psychopathy. *Journal of Interpersonal Violence, 13,* 514–527.

Heinze, M. C., & Grisso, T. (1996). Review of instruments assessing parenting competencies used in child custody evaluations. *Behavioral Sciences and the Law, 14,* 293–313.

Henning-Stout, M., & Brown-Cheatam, M. (1999). School psychology in a diverse world: Considerations for practice, research and training. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 1041–1055). New York: Wiley.

Herr, E. L., & Cramer, S. H. (1996). *Career guidance and counseling through the lifespan* (5th ed.). New York: HarperCollins.

Hobbs, N. (1951). *Report of the American Psychological Association's Committee on Ethical Standards for Psychology to the Council of Representatives*. Washington, DC: American Psychological Association.

Holland, J. L. (1976). Vocational preferences. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 521–570). New York: Rand McNally.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Hough, L. M. (1998). Personality at work: Issues and evidence. In M. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 131–166). Hillsdale, NJ: Erlbaum.

Howard, A. H. (Ed.). (1995). *The changing nature of work*. San Francisco: Jossey-Bass.

Hughes, J. N. (1999). Child psychotherapy. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 745–763). New York: Wiley.

Illback, R. J., Zins, J. E., & Maher, C. A. (1999). Program planning and evaluation: Principles, procedures and planned change. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 907–932). New York: Wiley.

International Test Commission. (2000). *International guidelines for test-use: Version 2000.* Retrieved April 25, 2000, from the World Wide Web: http://cwis.kub.nl/~fsw_1/itc/itcv2000.htm

Ivnik, R. J., Smith, G. E., Lucas, J. A., Petersen, R. C., Boeve, B. F., Kokmen, E., & Tangalos, E. G. (1999). Testing normal older persons three to four times at 1- to 2-year intervals: Defining normal variance. *Neuropsychology, 13,* 121–127.

Ivnik, R. J., Smith, G. E., Petersen, R. C., Boeve, B. F., Kokmen, E., & Tangalos, E. G. (in press). Diagnostic accuracy of four approaches to interpreting neuropsychological test data. *Neuropsychology*.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19.

Jacob-Timm, S., & Hartshorne, T. (1998). *Ethics and law for school psychologists* (2nd ed.). Brandon, VT: Clinical Psychology.

Kamphaus, R. W., Reynolds, C. R., & Imperato-McCammon, C. (1999). Roles of diagnosis and classification in school psychology. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 292–306). New York: Wiley.

Kane, J. S. (1986). Performance distribution assessment. In R. Berk (Ed.), *Performance assessment: Methods and applications* (pp. 237–273). Baltimore: Johns Hopkins University Press.

Kapes, J. T., Mastie, M. M., & Whitfield, E. A. (1994). *A counselor's guide to career assessment instruments*. Alexandria, VA: National Career Development Association.

Kilpatrick, D. G., & Resnick, H. S. (1993). Posttraumatic stress disorder associated with exposure to criminal victimization in clinical and community populations. In J. R. T. Davidson & E. B. Foa (Eds.), *Posttraumatic stress disorder: DSM–IV and beyond* (pp. 113–146). Washington, DC: American Psychiatric Press.

Knapp, J. E., & Knapp, L. G. (1995). Practice analysis: Building the foundation for validity. In J. C. Impara (Ed.), *Licensure testing: Purposes, procedures, and practices* (pp. 93–116). Lincoln, NE: Buros Institute of Mental Measurements.

Komaki, J., Collins, R. L., & Penn, P. (1982). The role of performance antecedents and consequences in work motivation. *Journal of Applied Psychology, 67,* 334–340.

Koocher, G. P. (1993). Ethical issues in the psychological assessment of children. In T. H. Ollendick & M. Hersen (Eds.), *Handbook of child and adolescent assessment: General psychology series* (Vol. 167, pp. 51–61). Boston: Allyn & Bacon.

Koocher, G. P., & Keith-Spiegel, P. C. (1998). *Ethics in psychology: Professional standards and cases* (2nd ed.). New York: Oxford University Press.

Koocher, G. P., Norcross, J. C., & Hill, S. S. III. (1998). *Psychologists' desk reference.* New York: Oxford University Press.

Kozlowski, S. W. J., Chao, G. T., Smith, E. M., & Hedlund, J. (1993). Organizational downsizing: Strategies, interventions, and research implications. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 263–332). Chichester, United Kingdom: Wiley.

Kranzler, J. H. (1999). Current contributions of the psychology of individual differences to school psychology. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 223–246). New York: Wiley.

Landy, F. J., & Farr, J. L. (1983). *The measurement of work performance*. New York: Academic Press.

Lemsky, C. M., Smith, G. E., Malec, J. F., & Ivnik, R. J. (1996). Identifying risk for functional impairment using cognitive measures: An application of CART modeling. *Neuropsychology, 10,* 368–375.

London, M. (1995). *Employees, careers, and job creation: Developing growth-oriented human resource strategies and programs*. San Francisco: Jossey-Bass.

Lowman, R. L. (1991). *The clinical practice of career assessment*. Washington, DC: American Psychological Association.

Lucas, L., & Epperson, D. (1988). Personality types in vocationally undecided students. *Journal of College Student Development, 29,* 460–466.

Lucas, M. S. (1992). Problems expressed by career and non-career help seekers: A comparison. *Journal of Counseling and Development, 70,* 417–420.

Macmann, G. M., & Barnett, D. (1999). Diagnostic decision making in school psychology: Understanding and coping with uncertainty. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 519–548). New York: Wiley.

Martzke, J. S., Swan, C. S., & Varney, N. R. (1991). Posttraumatic anosmia and orbital frontal damage: Neuropsychological and neuropsychiatric correlates. *Neuropsychology, 5,* 213–225.

Maruish, M. E. (1999). *The use of psychological testing for treatment planning and outcomes assessment* (2nd ed.). Mahwah, NJ: Erlbaum.

Medway, F. J., & Cafferty, T. P. (1999). Contributions of social psychology to school psychology. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 194–222). New York: Wiley.

Megargee, E. (1979). Development and validation of an MMPI-based system for classifying criminal offenders. In J. N. Butcher (Ed.), *New developments in the use of the MMPI* (pp. 303–324). Minneapolis: University of Minnesota Press.

Megargee, E. I. (1994). Using the Megargee MMPI-based classification system with MMPI–2s of male prison inmates. *Psychological Assessment, 6,* 337–344.

Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (1997). *Psychological evaluations for the courts* (2nd ed.). New York: Guilford Press.

Monahan, J., & Steadman, H. J. (1996). Violent storms and violent people: How meteorology can inform risk communication in mental health law. *American Psychologist, 51,* 931–938.

Moreland, K. L., Eyde, L. D., Robertson, G. J., Primoff, E. S., & Most, R. B. (1995). Assessment of test user qualifications: A research-based measurement procedure. *American Psychologist, 50,* 14–23.

Murphy, K. R. (Ed.). (1996). *Individual differences and behavior in organizations*. San Francisco: Jossey-Bass.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational and goal-based perspectives*. Thousand Hills, CA: Sage.

National Association of School Psychologists. (1992). *Standards for the provision of school psychological services*. Silver Spring, MD: Author.

National Council on Measurement in Education. (1995). *Code of professional responsibilities in educational measurement*. Washington, DC: Author.

Nuttall, E. V., Romero, I., & Kalesnik, J. (1999). *Assessing and screening preschoolers: Psychological and educational dimensions*. Boston: Allyn & Bacon.

Ogloff, J. R. P., Tomkins, A. J., & Bersoff, D. N. (1996). Education and training in psychology and law/criminal justice: Historical foundations, present structures, and future developments. *Criminal Justice and Behavior, 23,* 200–235.

Osipow, S. H., & Fitzgerald, L. F. (1996). *Theories of career development.* Boston: Allyn & Bacon.

Ostroff, C., & Ford, J. K. (1989). Assessing training needs: Critical levels of analysis. In I. L. Goldstein (Ed.), *Training and development in organizations* (pp. 25–62). San Francisco: Jossey-Bass.

Otto, R. (1992). Prediction of dangerous behavior: A review and analysis of "second generation" research. *Forensic Reports, 5,* 103–133.

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., & Fleishman, E. A. (Eds.). (1999). *An occupational information system for the 21st century: The development of O*NET*. Washington, DC: American Psychological Association.

Pickman, A. J. (1994). *The complete guide to outplacement counseling*. Hillsdale, NJ: Erlbaum.

Reschly, D. J., & Bersoff, D. N. (1999). Law and school psychology. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 1077–1112). New York: Wiley.

Reynolds, C. R. (1990). Conceptual and technical problems in learning disability diagnosis. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (pp. 571–592). New York: Guilford Press.

Reynolds, C. R., & Kamphaus, R. W. (Eds.). (1990). *Handbook of psychological and educational assessment of children.* New York: Guilford Press.

Robertson, G. J., & Eyde, L. D. (1986). Establishing test purchaser qualifications: Present practices and future needs. In R. B. Most (Ed.), *Test purchaser qualifications: Present practice, professional needs, and a proposed system. Issues in Scientific Psychology* (pp. 7–12). Washington, DC: American Psychological Association, Science Directorate.

Rogers, R. (1986). *Conducting insanity evaluations.* New York: Van Nostrand Reinhold.

Rogers, R. (1997). *Clinical assessment of malingering and deception.* New York: Guilford Press.

Sales, B., & Perrin, G. (1993). Artificial legal standards in mental–emotional injury legislation. *Behavioral Sciences and the Law, 11,* 193–203.

Salvia, J., & Ysseldyke, J. E. (1995). *Assessment.* Boston: Houghton Mifflin.

Sandoval, J., Frisby, C. L., Geisinger, K. F., Scheuneman, J. D., & Grenier, J. R. (Eds.). (1998). *Test interpretation and diversity: Achieving equity in assessment*. Washington, DC: American Psychological Association.

Sawrie, S. M., Chelune, G. J., Naugle, R. I., & Luders, H. O. (1996). Empirical methods for assessing meaningful neuropsychological change following epilepsy surgery. *Journal of the International Neuropsychological Society, 2,* 556–564.

Schmitt, K., & Shimberg, B. (1996). *Demystifying occupational and professional regulation: Answers to questions you may have been afraid to ask.* Lexington, KY: Council on Licensure, Enforcement and Regulation.

Schmitt, N., & Borman, W. C. (Eds.). (1993). *Personnel selection in organizations.* San Francisco: Jossey-Bass.

Shapiro, E. S., & Cole, C L. (1994). *Behavior change in the classroom: Self-management interventions.* New York: Guilford Press.

Shapiro, E. S., & Elliott, S. N. (1999). Curriculum-based assessment and other performance based assessment strategies. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 383–408). New York: Wiley.

Shuman, D. W., Greenberg, S. A., Heilbrun, K., & Foote, W. E. (1999). An immodest proposal: Should treating mental health professionals be barred from testifying about their patients? *Behavioral Sciences and the Law, 16,* 509–523.

Simner, M. L. (1994). *Draft of final report of the Professional Affairs Committee Working Group on Test Publishing Industry Safeguards.* Ottawa, Ontario, Canada: Canadian Psychological Association.

Swanson, J. W. (1994). Mental disorder, substance abuse and community violence: An epidemiological approach. In J. Monahan & H. J. Steadman (Eds.), *Violence and mental*

disorder: Developments in risk assessment (pp. 101–136). Chicago: University of Chicago Press.

Talge, J. von (1995). Overcoming courtroom challenges to the new *DSM–IV*: I. The

major changes in *DSM–IV*. *American Journal of Forensic Psychology, 13,* 5–29.

Tharinger, D. J., & Lambert, N. M. (1999). The application of developmental psychology

to school psychology: Informing assessment, intervention and prevention efforts. In C. R.

Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 132–166).

New York: Wiley.

Tinsley, H. E., & Barrett, T. C. (1977). Measuring vocational self-concept crystallization.

Journal of Vocational Behavior, 11, 305–313.

Tinsley, H. E., & Bradley, R. W. (1986). Test interpretation. *Journal of Counseling and

Development, 64,* 462–466.

Tyler, B. (1986). The use of tests by psychologists: Report on a survey of British

Psychological Society members. *Bulletin of the International Test Commission, 22,* 7–18.

Varney, N. R., & Menefee, L. (1993). Psychosocial and executive deficits following

closed head injury: Implications for orbital frontal cortex. *Journal of Head Trauma

Rehabilitation, 8,* 32–44.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). Assessing risk of

violence to others. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment,

and treatment* (pp. 251–277). New York: Guilford Press.

Weissman, H. N. (1991). Forensic psychological assessment and the effects of protracted

litigation on impairment in personal injury litigation. *Forensic Reports, 4,* 417–429.

White, L. (1987). The mental illness defense in the capital penalty hearing. *Behavioral Sciences and the Law, 5,* 397–410.

Woody, R. H., La Voie, J. C., & Epps, S. (1992). *School psychology: A developmental and social systems approach.* Boston: Allyn & Bacon.

Ysseldyke, J. E., & Elliott, J. (1999). Effective instructional practices: Implications for assessing instructional environments. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 497–518). New York: Wiley.

Zeidner, J., & Johnson, C. D. (1994). Is personnel classification a concept whose time has passed? In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 431–442). Hillsdale, NJ: Erlbaum.

Zunker, V. G. (1990). *Career counseling: Applied concepts in life planning* (3rd ed.). Pacific Grove, CA: Brooks/Cole.