

[Continue](#)



# The Building Resiliency Workbook

Facilitator Reproducible Self-Assessments, Exercises & Educational Handouts



Ester A. Leutenberg  
John J. Liptak, EdD

Illustrated by  
Amy L. Brodsky, LISW-S



Julie Racine PMHNP-BC  
Psychiatric Mental Health Nurse Practitioner

## AUTHORIZATION TO DISCLOSE CONFIDENTIAL INFORMATION

Name of Client: \_\_\_\_\_ DOB: \_\_\_\_\_ Client: \_\_\_\_\_

I hereby authorize Sweetser, its authorized employees and agents to: (Check all appropriate boxes)

Obtain written medical/clinical records:  Disclose written medical/clinical records:  Verbally discuss:

Organization/Individual: \_\_\_\_\_ Relationship: \_\_\_\_\_

Street Address: \_\_\_\_\_ Phone #: \_\_\_\_\_

City/State/Zip: \_\_\_\_\_ Fax #: \_\_\_\_\_

This authorization applies to all Sweetser programs that I am currently enrolled in or have been enrolled in unless program exclusion is noted: *DO NOT release information from following noted program(s):* \_\_\_\_\_

The medical/clinical records and information include the following: (mark with X only those items to be disclosed)

Treatment Plan  Psychological Testing  Discharge Summary  Psychiatric Evaluation   
 Progress Notes  Medical Information (including Labs, Med List)  Crisis Intervention Assessment  
 Diagnoses  Comprehensive Assessment  Other Records (be specific): \_\_\_\_\_

The records and information are to be provided or obtained for purposes of: (Place an X next to all appropriate responses)

Ongoing Treatment  Educational  Financial  To coordinate treatment efforts  
 Aftercare Treatment  Legal  Other (be specific): \_\_\_\_\_

1.  I DO  I DO NOT Authorize information which refers to treatment or diagnosis of alcohol or drug abuse to be disclosed or obtained.

**IMPORTANT:** If checked "I DO", then the client MUST sign this consent, regardless of age. I understand my alcohol and substance abuse record is protected under Federal Regulation, 42 CFR 2.101, which prohibits these records from being disclosed or re-disclosed without written consent, unless otherwise provided in Regulations. While the Federal Regulations protects information from being re-disclosed I understand that Sweetser can not guarantee that the recipient will not re-disclose this information to a third party.

2.  I DO  I DO NOT authorize information concerning diagnosis and treatment of mental health conditions to be disclosed or obtained.

3.  I DO  I DO NOT authorize information which refers to treatment or diagnosis of HIV infection or AIDS to be disclosed or obtained.

4.  I DO  I DO NOT wish to review written information prior to its being disclosed. (No X is needed for Verbal communications)

5.  I DO  I DO NOT want a copy of this consent.

The records and information are to be provided or obtained for purposes of: (Check all appropriate boxes)  Educational  
 Ongoing/ Aftercare Treatment  Financial  To coordinate treatment efforts  Legal  Other (be specific): \_\_\_\_\_

### I understand that

- I can refuse to disclose some or all of the information in my treatment records, but if I do so, it could result in an improper diagnosis or treatment, or a denial of coverage or of a claim for health benefits/other insurance, or other adverse consequences.
- Sweetser's provision of services does not depend on my giving this consent, except that my refusing consent connected with a research project may result in my not receiving treatment as a participant in that project.
- Any records and information disclosed to a recipient outside Sweetser may potentially be re-disclosed and no longer be protected by Federal or State law.
- I may revoke this authorization at any time either verbally or in writing. A revocation does not apply to any actions previously taken in reliance on my consent, including disclosures already made or services already rendered.

This consent is effective until \_\_\_\_\_ (Maximum is one year for mental health services, six months for children in residential care only, and ninety days for one-time disclosures).  I DO  I DO NOT authorize future disclosures regarding these records to the same individual's entity during this time period.

Signature of Client: \_\_\_\_\_ Date: \_\_\_\_\_

Signature of Parent/Guardian: \_\_\_\_\_ Date: \_\_\_\_\_

Printed Name: \_\_\_\_\_

Staff Signature and Title: \_\_\_\_\_ Date: \_\_\_\_\_

# Psychometric Properties of the 44-Item Version of Ryff's Psychological Well-Being Scale

Éva Kállay and Claudia Rus

Babes-Bolyai University, Cluj-Napoca, Romania

**Abstract.** This study examined the factorial validity and reliability ( $\Omega$ ) of the nonreversed, 44-item version of Ryff's Psychological Well-Being Scale (PWBS; Ryff, 1989) on a Romanian convenience sample of 664 participants from the general population. The results showed that the correlated six-factor model presented a relatively good fit,  $\chi^2(987) = 2922.85, p < .001, RMSEA = .059, RMSEA 90\% CI = [0.56, 0.62], SRMR = .048, CFI = .973$ , compared to single-factor and independent six-factor models. Based on the ACFI value, we found no significant differences between the correlated six-factor and the hierarchical model. Although the correlated six-factor model had a relatively good fit, the high correlations between the six latent factors suggest a high overlap among them. Our results indicate that well-being can be conceptualized as a second-order factor encompassing six dimensions, represented by autonomy, positive relations, environmental mastery, personal growth, purpose in life, and self-acceptance. The value of the  $\Omega$  reliability coefficient of the six subscales as well as the whole instrument was above .70. The present study has a practical implication by highlighting the factorial validity of a shorter (44-item) instrument, thus shortening the time necessary for data collection.

**Keywords:** psychological well-being, factorial validity, omega-weighted reliability, psychometrics, confirmatory factor analysis

## Introduction

Until quite recently, psychological research almost exclusively focused on the investigation of negative human functioning (Huppert, 2010). But if we take into consideration that health is defined as a state of complete physical, mental, and social well-being, i.e., more than just the absence of illness and malfunctioning (WHO, 1948), then the promotion of well-being becomes imperative (Kirkwood, Bond, May, McKeith, & Teh, 2010). Since the dimensions of human functioning are strongly related and interwoven, having a direction and a goal in life, developing and maintaining positive, high-quality human relationships as well as striving toward attaining one's potential may all significantly contribute to the maintenance of health and to recovery from illness (Ryff & Singer, 1998). As Ryff, Singer, and Love (2004) emphasized, "the experience of well-being contributes to the effective function of multiple biological systems, which may help keep the organism from succumbing to disease, or when illness or adversity occurs, may help promote rapid recovery" (p. 1383).

The promotion of well-being becomes even more important if we consider the rapid social and economic changes occurring world-wide. Globalization, demographic shifts, increasing competition and pressure for excellence in productivity, less predictable career paths, increased uncer-

tainty at the workplace, changes in basic norms and values systems, etc. (Amundson, 2006), have not only impacted our working lives, but have also carried over their effects to the domain of personal lives (Wechslein, 2008). These changes were also experienced in Romania during the last decade. Depending on the constant interplay of a myriad of factors, successful adaptation would lead to flourishing (Seligman, 2011). On the other hand, unsuccessful adaptation would lead to emptiness and stagnation, elevated levels of negative emotions, low levels of positive emotions, also known as languishing (Keyes, 2008), while severe maladaptation leads to different forms of physical and psychological disorders (Levenson, 2005). Thus, the last couple of decades have witnessed a significant increase of interest in the investigation of well-being (Ryff & Singer, 2006), targeting the identification of its constituents, and mapping its characteristics, causes, and consequences (Huppert, Keverne, & Bayliss, 2004).

From its inception, the empirical investigation of well-being had two separate thrusts, concentrating on these two distinct approaches: research on hedonic well-being (subjective well-being, SWB) and eudaimonic well-being (psychological well-being, PWB) (Ryan & Deci, 2001). In the hedonic approach, a well-lived life is equated with happiness, contentment, and life satisfaction, as experienced on the subjective level of human functioning (see Diener,

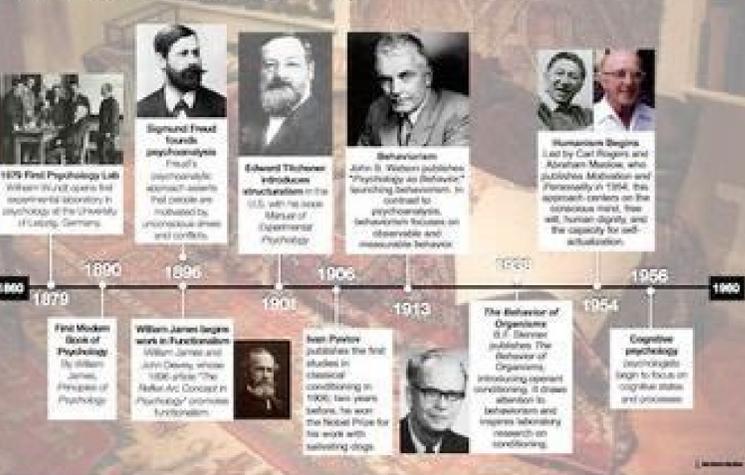
© 2013 Hogrefe Publishing

European Journal of Psychological Assessment 2014, Vol. 30(1), 15–21

DOI: 10.1027/1015-5759/a000163

## THE HISTORY OF PSYCHOLOGY

### 10 Years: A Concise History



Test validity is the extent to which a test (such as a chemical, physical, or scholastic test) accurately measures what it is supposed to measure. In the fields of psychological testing and educational testing, "validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (1). Although classical models divided the concept into various "validities" (such as content validity, criterion validity, and construct validity), (2) the currently dominant view is that validity is a single unitary construct. (3) Validity is generally considered the most important issue in psychological and educational testing (4) because it concerns the meaning placed on test results. (5) Though many textbooks present validity as a static construct, (5) various models of validity have evolved since the first published recommendations for constructing psychological and education tests. (6) These models can be categorized into two primary groups: classical models, which include several types of validity, and modern models, which present validity as a single construct. The modern models reorganize classical "validities" into either "aspects" of validity (3) or "types" of validity-supporting evidence. (1) Test validity can itself be tested/validated using tests of inter-rater reliability, intra-rater reliability, repeatability (test-retest reliability), and other traits, usually via multiple runs of the test whose results are compared. Statistical analysis helps determine whether the differences between the various results either are large enough to be a problem or are acceptably small. Historical background Although psychologists and educators were aware of several facets of validity before World War II, their methods for establishing validity were commonly restricted to correlations of test scores with some known criterion. (7) Under the direction of Lee Cronbach, the 1954 Technical Recommendations for Psychological Tests and Diagnostic Techniques (6) attempted to clarify and broaden the scope of validity by dividing it into four parts: (a) concurrent validity, (b) predictive validity, (c) content validity, and (d) construct validity. Cronbach and Meehl's subsequent publication (8) grouped predictive and concurrent validity into a "criterion-orientation", which eventually became criterion validity. Over the next four decades, many theorists, including Cronbach himself, (9) voiced their dissatisfaction with this three-in-one model of validity. (10) [1] [12] Their arguments culminated in Samuel Messick's 1995 article that described validity as a single construct, composed of six "aspects". (3) In his view, various inferences made from test scores may require different types of evidence, but not different validities. The 1999 Standards for Educational and Psychological Testing (1) largely codified Messick's model. They describe five types of validity-supporting evidence that incorporate each of Messick's aspects, and make no mention of the classical models' content, criterion, and construct validities. Validation process According to the 1999 Standards, (1) validation is the process of gathering evidence to provide "a sound scientific basis" for interpreting the scores as proposed by the test developer and/or the test user. Validation therefore begins with a framework that defines the scope and aspects (in the case of multi-dimensional scales) of the proposed interpretation. The framework also includes a rational justification linking the interpretation to the test in question. Validity researchers then list a series of propositions that must be met if the interpretation is to be valid. Or, conversely, they may compile a list of issues that may threaten the validity of the interpretations. In either case, the researchers proceed by gathering evidence – be it original empirical research, meta-analysis or review of existing literature, or logical analysis of the issues – to support or to question the interpretation's propositions (or the threats to the interpretation's validity). Emphasis is placed on quality, rather than quantity, of the evidence. A single interpretation of any test result may require several propositions to be true (or may be questioned by any one of a set of threats to its validity). Strong evidence in support of a single proposition does not lessen the requirement to support the other propositions. Evidence to support (or question) the validity of an interpretation can be categorized into one of five categories: Evidence based on test content Evidence based on response processes Evidence based on internal structure Evidence based on relations to other variables Evidence based on consequences of testing Techniques to gather each type of evidence should only be employed when they yield information that would support or question the propositions required for the interpretation in question. Each piece of evidence is finally integrated into a (Cronbach, 2009; Weiner, 2003). Such information is obtained through a variety of methods and measures, with relevant sources determined by the specific purposes of the evaluation. Sources of information may include: Records (e.g., medical, educational, occupational, legal) obtained from the referral source; Records obtained from other organizations and agencies that have been identified as potentially relevant; Interviews conducted with the person being examined; Behavioral observations; Interviews with corroborative sources such as family members, friends, teachers, and others; and Formal psychological or neuropsychological testing. Agreements across multiple measures and sources, as well as discrepant information, enable the creation of a more comprehensive understanding of the individual being assessed, ultimately leading to more accurate and appropriate clinical conclusions (e.g., diagnosis, recommendations for treatment planning). The clinical interview remains the foundation of many psychological and neuropsychological assessments. Interviewing may be structured, semistructured, or open in nature, but the goal of the interview remains consistent—to identify the nature of the client's presenting issues, to obtain direct historical information from the examinee regarding such concerns, and to explore historical variables that may be related to the complaints being presented. In addition, the interview element of the assessment process allows for behavioral observations that may be useful in describing the client, as well as discerning the convergence with known diagnoses. Based on the information and observations gained in the interview, assessment instruments may be selected, corroborative informants identified, and other historical records recognized that may aid the clinician in reaching a diagnosis. Conceptually, clinical interviewing explores the presenting complaint(s) (i.e., referral question), informs the understanding of the case history, aids in the development of hypotheses to be examined in the assessment process, and assists in determination of methods to address the hypotheses through formal testing. An important piece of the assessment process and the focus of this report, psychological testing consists of the administration of one or more standardized procedures under particular environmental conditions (e.g., quiet, good lighting) in order to obtain a representative sample of behavior. Such formal psychological testing may involve the administration of interviews, questionnaires, surveys, and/or tests, selected with regard to the specific examinee and his or her circumstances, that offer information to respond to an assessment question. Assessments, then, serve to respond to questions through the use of tests and other procedures. It is important to note that the selection of appropriate tests requires an understanding of the specific circumstances of the individual being assessed, falling under the purview of clinical judgment. For this reason, the committee refrains from recommending the use of any specific test in this report. Any reference to a specific test is to provide an illustrative example, and should not be interpreted as an endorsement by the committee for use in any specific situation; such a determination is best left to a qualified assessor familiar with the specific circumstances surrounding the assessment. To respond to questions regarding the use of psychological tests for the assessment of the presence and severity of disability due to mental disorders, this chapter provides an introductory review of psychological testing. The chapter is divided into three sections: (1) types of psychological tests, (2) psychometric properties of tests, and (3) test user qualifications and administration of tests. Where possible an effort has been made to address the context of disability determination; however, the chapter is primarily an introduction to psychological testing. There are many facets to the categorization of psychological tests, and even more if one includes educationally oriented tests; indeed, it is often difficult to differentiate many kinds of tests as purely psychological tests as opposed to educational tests. The ensuing discussion lays out some of the distinctions among such tests; however, it is important to note that there is no one correct cataloging of the types of tests because the different categorizations often overlap. Psychological tests can be categorized by the very nature of the behavior they assess (what they measure), their administration, their scoring, and how they are used. Figure 3-1 illustrates the types of psychological measures as described in this report. One of the most common distinctions made among tests relates to whether they are measures of typical behavior (often non-cognitive measures) versus tests of maximal performance (often cognitive tests) (Cronbach, 1949). As a measure of typical behavior, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures. There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of ability and tests of achievement; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences (Cronbach, 1949). As a measure of ability, those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as non-cognitive measures. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as cognitive tests. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report. One distinction among non-cognitive measures is whether the stimuli composing the measure are structured or unstructured. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it

intended purposes; and3. Criterion-related evidence of validity. The degree to which the test's score correlates with other measurable, reliable, and relevant variables (i.e., criterion) thought to measure the same construct.6. Internal validity. The degree to which test scores represent everyday levels of functioning (e.g., impact of disability on an individual's ability to function independently).3. Cultural validity: The degree to which test content and procedures accurately reflect the sociocultural context of the subjects being tested.Each of these forms of validity poses complex questions regarding the use of particular psychological measures with the SSA population. For example, ecological validity is especially critical in the use of psychological tests with SSA given that the focus of the assessment is on examining everyday levels of functioning. Measures like intelligence tests have been sometimes criticized for lacking ecological validity (Groth-Marnat, 2009; Groth-Marnat and Teal, 2000). Alternatively, "research suggests that many neuropsychological tests have a moderate level of ecological validity when predicting everyday cognitive functioning" (Chaytor and Schmitter-Edgecombe, 2003, p. 181).More recent discussions on validity have shifted toward an argument-based approach to validity, using a variety of evidence to build a case for validity of test score interpretation (Furr and Bacharach, 2013). In this approach, construct validity is viewed as an overarching paradigm under which evidence is gathered from multiple sources to build a case for validity of test score interpretation. Five key sources of validity evidence that affect the degree to which a test fulfills its purpose are generally considered (AERA et al., 2014; Furr and Bacharach, 2013; Sireci and Sukin, 2013):1. Test content: Does the test content reflect the important facets of the construct being measured? Are the test items relevant and appropriate for measuring the construct and congruent with the purpose of testing?2. Relation to other variables: Is there a relationship between test scores and other criterion or constructs that are expected to be related?3. Internal structure: Does the actual structure of the test match the theoretically based structure of the construct?4. Response processes: Are respondents applying the theoretical constructs or processes the test is designed to measure?5. Consequences of testing: What are the intended and unintended consequences of testing? As part of the development of any psychometrically sound measure, explicit methods and procedures by which tasks should be administered are determined and clearly spelled out. This is what is commonly known as standardization. Typical standardized administration procedures or expectations include (1) a quiet, relatively distraction-free environment, (2) precise reading of scripted instructions, and (3) provision of necessary tools or stimuli. All examinees use such methods and procedures during the process of collecting the normative data, and such procedures normally should be used in any other administration, which enables application of normative data to the individual being evaluated (Lezak et al., 2012).Standardized tests provide a set of normative data (i.e., norms), or scores derived from groups of people for whom the measure is designed (i.e., the designated population) to which an individual's performance can be compared. Norms consist of transformed scores such as percentiles, cumulative percentiles, and standard scores (e.g., T-scores, Z-scores, stanines, IQs), allowing for comparison of an individual's test results with the designated population. Without standardized administration, the individual's performance may not accurately reflect his or her ability. For example, an individual's abilities may be overestimated if the examiner provides additional information or guidance than what is outlined in the test administration manual. Conversely, a claimant's abilities may be underestimated if appropriate instructions, examples, or prompts are not presented. When nonstandardized administration techniques must be used, norms should be used with caution due to the systematic error that may be introduced into the testing process; this topic is discussed in detail later in the chapter. It is important to clearly understand the population for which a particular test is intended. The standardization sample is another name for the norm group. Norms enable one to make meaningful interpretations of obtained test scores, such as making predictions based on evidence. Developing appropriate norms depends on size and representativeness of the sample. In general, the more people in the norm group the closer the approximation to a population distribution so long as they represent the group who will be taking the test.Norms should be based upon representative samples of individuals from the intended test population, as each person should have an equal chance of being in the standardization sample. Stratified samples enable the test developer to identify particular demographic characteristics represented in the population and more closely approximate these features in proportion to the population. For example, intelligence test scores are often established based upon census-based norming with proportional representation of demographic features including race and ethnic group membership, parental education, socioeconomic status, and geographic region of the country.When tests are applied to individuals for whom the test was not intended and, hence, were not included as part of the norm group, inaccurate scores and subsequent misinterpretations may result. Tests administered to persons with disabilities often raise complex issues. Test users sometimes use psychological tests that were not developed or normed for individuals with disabilities. It is critical that tests used with such persons (including SSA disability claimants) include attention to representative norming samples; when such norming samples are not available, it is important for the assessor to note that the test or tests used are not based on representative norming samples and the potential implications for interpretation (Turner et al., 2001).Performance on psychological tests often has significant implications (high stakes) in our society. Tests are in part the gatekeepers for educational and occupational opportunities and play a role in SSA determinations. As such, results of psychological testing may have positive or negative consequences for an individual. Often such consequences are intended; however, there is the possibility for unintended negative consequences. It is imperative that issues of test fairness be addressed so no individual or group is disadvantaged in the testing process based upon factors unrelated to the areas measured by the test. Biases simply cannot be present in these kinds of professional determinations. Moreover, it is imperative that research demonstrates that measures can be fairly and equitably used with members of the various subgroups in our population. It is important to note that there are people from many language and cultural groups for whom there are no available tests with norms that are appropriately representative for them. As noted above, in such cases it is important for assessors to include a statement about this situation whenever it applies and potential implications on scores and resultant interpretation.While all tests reflect what is valued within a particular cultural context (i.e., cultural loading), bias refers to the presence of systematic error in the measurement of a psychological construct. Bias leads to inaccurate test results given that scores reflect either overestimations or underestimations of what is being measured. When bias occurs based upon culturally related variables (e.g., race, ethnicity, social class, gender, educational level) then there is evidence of cultural test bias (Suzuki et al., 2014).Relevant considerations pertain to issues of equivalence in psychological testing as characterized by the following (Suzuki et al., 2014, p. 260):1. Functional: Whether the construct being measured occurs with equal frequency across groups;2. Conceptual: Whether the item information is familiar across groups and means the same thing in various cultures;3. Scalar: Whether average score differences reflect the same degree, intensity, or magnitude for different cultural groups;4. Linguistic: Whether the language used has similar meaning across groups; and5. Metric: Whether the scale measures the same behavioral qualities or characteristics and the measure has similar psychometric properties in different cultures.It must be established that the measure is operating appropriately in various cultural contexts. Test developers address issues of equivalence through procedures includingExpert panel reviews (i.e., professionals review item content and provide informed judgments regarding potential biases);Examination of differential item functioning (DIF) among groups;Statistical procedures allowing comparison of psychometric features of the test (e.g., reliability coefficients) based on different population samples;Exploratory and confirmatory factor analysis; structural equation modeling (i.e., examination of the similarities and differences of the constructs structure), and measurement invariance; andMean score differences taking into consideration the spread of scores within particular racial and ethnic groups as well as among groups.Cultural equivalence refers to whether "interpretations of psychological measurements, assessments, and observations are similar if not equal across different ethnocultural populations" (Trimble, 2010, p. 316). Cultural equivalence is a higher order form of equivalence that is dependent on measures meeting specific criteria indicating that a measure may be appropriately used with other cultural groups beyond the one for which it was originally developed. Trimble (2010) notes that there may be upward of 50 or more types of equivalence that affect interpretive and procedural practices in order to establish cultural equivalence. For most of the 20th century, the dominant measurement model was called classical test theory. This model was based on the notion that all scores were composed of two components: true score and error. One can imagine a "true score" as a hypothetical value that would represent a person's actual score were there no error present in the assessment (and unfortunately, there is always some error, both random and systematic). The model further assumes that all error is random and that any correlation between error and some other variable, such as true scores, is effectively zero (Geisinger, 2013). The approach leans heavily on reliability theory, which is largely derived from the premises mentioned above.Since the 1950s and largely since the 1970s, a newer mathematically sophisticated model developed called item response theory (IRT). The premise of these IRT models is most easily understood in the context of cognitive tests, where there is a correct answer to questions. The simplest IRT model is based on the notion that the answering of a question is generally based on only two factors: the difficulty of the question and the ability level of the test-taker. Computer-adaptive testing estimates scores of the test-taker after each response to a question and adjusts the administration of the next question accordingly. For example, if a test-taker answers a question correctly, he or she is likely to receive a more difficult question next. If one, on the other hand, answers incorrectly, he or she is more likely to receive an easier question, with the "running score" held by the computer adjusted accordingly. It has been found that such computer-adaptive tests can be very efficient.IRT models have made the quality of test forms far easier. Equating tests permits one to use different forms of the same examination with different test items to yield fully comparable scores due to slightly different item difficulties across forms. To convert the values of items difficulty to determine the test-taker's ability scores one needs to have some common items across various tests; these common items are known as anchor items. Using such items, one can essentially establish a fixed reference score from other groups on these values.As noted above, there are a number of common IRT models. Among the most common are the one-, two-, and three-parameter models. The one-parameter model is the one already described; the only item parameter is item difficulty. A two-parameter model adds a second parameter to the first, related to item discrimination. Item discrimination is the ability of the item to differentiate those lacking the ability in high degree from those holding it. This two-parameter models are often used for tests like essay tests where one cannot achieve a high score by guessing or using other means to answer correctly. The three-parameter IRT model contains a third parameter, that factor related to chance level correct scoring. This parameter is sometimes called the pseudo-guessing parameter, and this model is generally used for large-scale multiple-choice testing programs. These models, because of their lessened reliance on the sampling of test-takers, are very useful in the equating of tests that is the setting of scores to be equivalent regardless of the form of the test one takes. In some high-stakes admissions tests such as the GRE, MCAT, and GMAT, for example, forms are scored and equated by virtue of IRT methods, which can perform such operations more efficiently and accurately than can be done with classical statistics. The test user is generally considered the person responsible for appropriate use of psychological tests, including selection, administration, interpretation, and use of results (AERA et al., 2014). Test user qualifications include attention to the purchase of psychological measures that specify levels of training, educational degree, areas of knowledge within domain of assessment (e.g., ethical administration, scoring, and interpretation of clinical assessment), certifications, licensure, and membership in professional organizations. Test user qualifications require psychometric knowledge and skills as well as training regarding the response use of tests (e.g., ethics), in particular, psychometric knowledge and skills, reliability and measurement error, validity and the meaning of test scores, normative interpretation of test scores, selection of appropriate tests, and test administration procedures. In addition, test user guidelines highlight the importance of understanding the impact of ethnic, racial, cultural, gender, age, educational, and linguistic characteristics in the selection and use of psychological tests (Turner et al., 2001).Test publishers provide detailed manuals regarding the operational definitions of the construct being assessed, norming sample, reading level of test items, completion time, administration, and scoring and interpretation of test scores. Directions presented to the examinee are provided verbatim and sample responses are often provided to assist the examiner in determining a right or wrong response or in awarding numbers of points to a particular answer. Ethical and legal knowledge regarding assessment competencies, confidentiality of test information, test security, and legal rights of test-takers are imperative. Resources like the Mental Measurements yearbook (MMY) provide descriptive information and evaluative reviews of commercially available tests to promote and encourage informed test selection (Buros, 2015). To be included, tests must contain sufficient documentation regarding their psychometric quality (e.g., validity, reliability, norming).In accordance with the Standards for Educational and Psychological Testing (AERA et al., 2014) and the APA's Guidelines for Test User Qualifications (Turner et al., 2001), many publishers of psychological tests employ a tiered system of qualification levels (generally A, B, C) required for the purchase, administration, and interpretation of such tests (e.g., PAR, n.d.; Pearson Education, 2015). Many instruments, such as those discussed throughout this report, would be considered qualification level C assessment methods, generally requiring an advanced degree, specialized psychometric and measurement knowledge, and formal training in administration, scoring, and interpretation. However, some may have less stringent requirements, for example, a bachelor's or master's degree in a related field and specialized training in psychometric assessment (often classified level B), or no special requirements (often classified level A) for purchase and use. While such categories serve as a general guide for necessary qualifications, individual test manuals provide additional details and specific qualifications necessary for administration, scoring, and interpretation of the test or measure. Given the need for the use of standardized procedures, any person administering cognitive or neuropsychological measures must be well trained in standardized administration protocols. He or she should possess the interpersonal skills necessary to build rapport with the individual being tested in order to foster cooperation and maximal effort during testing. Additionally, individuals administering tests should understand important psychometric properties, including validity and reliability, as well as factors that could emerge during testing to place either at risk. Many doctoral-level psychologists are well trained in test administration; in general, psychologists from clinical, counseling, school, or educational graduate psychology programs receive training in psychological test administration. For cases in which cognitive deficits are being evaluated, a neuropsychologist may be needed to most accurately evaluate cognitive functioning (see Chapter 5 for a more detailed discussion on administration and interpretation of cognitive tests). The use of non-doctoral-level psychometrists or technicians in psychological and neuropsychological test administration and scoring is also a widely accepted standard of practice (APA, 2010; Brandt and van Gorp, 1999; Pearson Education, 2015). Psychometrists are often bachelor's- or master's-level individuals who have received additional specialized training in standardized test administration and scoring. They do not practice independently or interpret test scores, but rather work under the close supervision and direction of doctoral-level clinical psychologists or neuropsychologists.Interpretation of testing results requires a higher degree of clinical training than administration alone. Threats to the validity of any psychological measure of a self-report nature oblige the test interpreter to understand the test and principles of test construction. In fact, interpreting tests results without such knowledge would violate the ethics code established for the profession of psychology (APA, 2010). SSA requires psychological testing be "individually administered by a qualified specialist ... currently licensed or certified in the state to administer, score, and interpret psychological tests and have the training and experience to perform the tests" (SSA, n.d.). Most doctoral-level clinical psychologists who have been trained in psychometric test administration are also trained in test interpretation. SSA (n.d.) also requires individuals who administer more specific cognitive or neuropsychological evaluations "be properly trained in this area of neuroscience." As such, clinical neuropsychologists—individuals who have been specifically trained to interpret testing results within the framework of brain-behavior relationships and who have achieved certain educational and training benchmarks as delineated by national professional organizations—may be required to interpret tests of a cognitive nature (AACN, 2007; NAN, 2001).Modification of procedures, including the use of interpreters and the administration of nonstandardized assessment procedures, may pose unique challenges to the psychologist by potentially introducing systematic error into the testing process. Such errors may be related to language, the use of translators, or examinee abilities (e.g., sensory, perceptual, and/or motor capacity). For example, if one uses a language interpreter, the potential for mistranslation may yield inaccurate scores. Use of translators is a nonpreferred option, and assessors need to be familiar with both the language and culture from which an individual comes to properly interpret test results, or even infer whether specific measures are appropriate. The adaptation of tests has become big business for testing companies, and many tests, most often measures developed in English for use in the United States, are being adapted for use in other countries. Such measures require changes in language, but translators must also be knowledgeable about culture and the environment of the region from which a person comes (ITC, 2005).For sensory, perceptual, or motor abilities, one may be altering the construct that the test is designed to measure. In both of these examples, one could be obtaining scores for which there is no referenced normative group to allow for accurate interpretation of results. While a thorough discussion of these concepts is beyond the scope of this report and is presented elsewhere, it may be stated that when a test is administered following a procedure that is outside of that which has been developed in the standardization process, conclusions drawn must recognize the potential for error in their creation.As noted in Chapter 2, SSA indicates that objective medical evidence may include the results of standardized psychological tests. Given the great variety of psychological tests, some are more objective than others. Whether a psychological test is appropriately considered objective has much to do with the process of scoring. For example, unstructured measures that call for open-ended responding rely on professional judgment and interpretation in scoring; thus, such measures are considered less than objective. In contrast, standardized psychological tests and measures, such as those discussed in the ensuing chapters, are structured and objectively scored. In the case of non-cognitive self-report measures, the respondent generally answers questions regarding typical behavior by choosing from a set of predetermined answers. With cognitive tests, the respondent answers questions or solves problems, which usually have correct answers, as well as he or she possibly can. Such measures generally provide a set of normative data (i.e., norms), or scores derived from groups of people for whom the measure is designed (i.e., the designated population), to which an individual's responses or performance can be compared. Therefore, standardized psychological tests and measures rely less on clinical judgment and are considered to be more objective than those that depend on subjective scoring. Unlike measurements such as weight or blood pressure standardized psychological tests require the individual's cooperation with respect to self-report or performance on a task. The inclusion of validity testing, which will be discussed further in Chapters 4 and 5, in the test or test battery allows for greater confidence in the test results. Standardized psychological tests that are appropriately administered and interpreted can be considered objective evidence.The use of psychological tests in disability determinations has critical implications for clients. As noted earlier, issues surrounding ecological validity (i.e., whether test performance accurately reflects real-world behavior) is of primary importance in SSA determination. Two approaches have been identified in relation to the ecological validity of neuropsychological assessment. The first focuses on "how well the test captures the essence of everyday cognitive skills" in order to "identify people who have difficulty performing real-world tasks, regardless of the etiology of the problem" (i.e., verisimilitude), and the second "relates performance on traditional neuropsychological tests to measures of real-world functioning, such as employment status, questionnaires, or clinician ratings" (i.e., veridicality) (Chaytor and Schmitter-Edgecombe, 2003, pp. 182–183). Establishing ecological validity is a complicated endeavor given the potential effect of non-cognitive factors (e.g., emotional, physical, and environmental) on test and everyday performance. Specific concerns regarding test performance include (1) the test environment is often not representative (i.e., artificial), (2) testing yields only samples of behavior that may fluctuate depending on context, and (3) clients may possess compensatory strategies that are not employable during the testing situation; therefore, obtained scores underestimate the test-taker's abilities.Activities of daily living (ADLs) and the client's likelihood of returning to work are important considerations in disability determinations. Occupational status, however, is complex and often multidetermined requiring that psychological test data be complemented with other sources of information in the evaluation process (e.g., observation, informant ratings, environmental assessments) (Chaytor and Schmitter-Edgecombe, 2003). Table 3-1 highlights major mental disorders, relevant types of psychological measures, and domains of functioning.Determination of disability is dependent on two key factors: the existence of a medically determinable impairment and associated limitations on functioning. As discussed in detail in Chapter 2, applications for disability follow a five-step sequential disability determination process. At Step 3 in the process, the applicant's reported impairments are evaluated to determine whether they meet or equal the medical criteria codified in SSA's Listing of Impairments. This includes specific symptoms, signs, and laboratory findings that substantiate the existence of an impairment (i.e., Paragraph A criteria) and evidence of associated functional limitations (i.e., Paragraph B criteria). If an applicant's impairments meet or equal the listing criteria, the claim is allowed. If not, residual functional capacity, including mental residual functional capacity, is assessed. This includes whether the applicant has the capacity for past work (Step 4) or any work in the national economy (Step 5). SSA uses a standard assessment that examines functioning in four domains: understanding and memory, sustained concentration and persistence, social interaction, and adaptation. Psychological testing may play a key role in understanding a client's functioning in each of these areas. Box 3-1 describes ways in which these four areas of core mental residual functional capacity are assessed ecologically. Psychological assessments often address these areas in a more structured manner through interviews, standardized measures, checklists, observations, and other assessment procedures.Descriptions of Tests by Four Areas of Core Mental Residual Functional Capacity. Remember location and work-like procedures Understand and remember very short and simple instructions This chapter has identified some of the basic foundations underlying the use of psychological tests including basic psychometric principles and issues regarding test fairness. Applications of tests can inform disability determinations. The next two chapters build on this overview, examining the types of psychological tests that may be useful in this process, including a review of selected individual tests that have been developed for measuring validity of presentation. Chapter 4 focuses on non-cognitive, self-report measures and symptom validity tests. Chapter 5 then focuses on cognitive tests and associated performance validity tests. Strengths and limitations of various instruments are offered, in order to subsequently explore the relevance for different types of tests for different claims, per category of disorder, with a focus on establishing the validity of the client's claim.REFERENCES AACN (American Academy of Clinical Neuropsychology). AACN practice guidelines for neuropsychological assessment and consultation. Clinical Neuropsychology. 2007;21(2):209–231. [PubMed: 17455014]AERA (American Educational Research Association). APA (American Psychological Association), and NCME (National Council on Measurement in Education). Standards for educational and psychological testing. Washington, DC: AERA; 2014. Brandt J, van Gorp W. American Academy of Clinical Neuropsychology policy on the use of non-doctoral-level personnel in conducting clinical neuropsychological evaluations. The Clinical Neuropsychologist. 1999;13(4):385–385. Chaytor N, Schmitter-Edgecombe M. The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. Neuropsychology Review. 2003;13(4):181–197. [PubMed: 15000225]Cronbach LJ. Essentials of psychological testing. New York: Harper; 1949. Cronbach LJ. Essentials of psychological testing. 2nd. Oxford, England: Harper; 1960. De Ayala RJ. Theory and practice of item response theory. New York: Guilford Publications; 2009. DeMars C. Item response theory. New York: Oxford University Press; 2010. Furr RM, Bacharach VR. Psychometrics: An introduction. Thousand Oaks, CA: Sage Publications, Inc.; 2013. Geisinger KF. Reliability. Geisinger KF, Bracken BA, Carlson JF, Hansen JC, Kuncel NR, Reise SP, Rodriguez MC, editors. Washington, DC: APA; APA handbook of testing and assessment in psychology. 2013;1:Groth-Marnat G. Handbook of psychological assessment. Hoboken, NJ: John Wiley & Sons; 2009. Groth-Marnat G, Teal M. Block design as a measure of everyday spatial ability: A study of ecological validity. Perceptual and Motor Skills. 2000;90(2):522–526. [PubMed: 10833749]Hambleton RK, Pitoniak MJ. Setting performance standards. Educational Measurement. 2006;4:433–470.ITC (International Test Commission). ITC guidelines for translating and adapting tests. Geneva, Switzerland: ITC; 2005. Lezak M, Howieson D, Bigler E, Tranel D. Neuropsychological assessment. 5th. New York: Oxford University Press; 2012. Sattler JM. Foundations of behavioral, social, and clinical assessment of children. 6th. La Mesa, CA: Jerome M. Sattler, Publisher, Inc.; 2014. Sireci SG, Sukin T. Test validity. Geisinger KF, Bracken BA, Carlson JF, Hansen JC, Kuncel NR, Reise SP, Rodriguez MC, editors. Washington, DC: APA; APA handbook of testing and assessment in psychology. 2013;1:Suzuki LA, Nagvi S, Hill JS. Assessing intelligence in a cultural context. Leong FT, Comas-Diaz L, Nagayama Hall GC, McLeod VC, Trimble JE, editors. Washington, DC: APA; APA handbook of multicultural psychology. 2014;1:Trimble JE. Encyclopedia of cross-cultural school psychology. New York: Springer; 2010. Cultural measurement equivalence, pp. 316–318.Turner SM, DeMers ST, Fox HR, Reed G. APA's guidelines for test user qualifications: An executive summary. American Psychologist. 2001;56(12):1099. Weiner IB. The assessment process. In: Weiner IB, editor. Handbook of psychology. Hoboken, NJ: John Wiley & Sons; 2003. 1This may be in comparison to a nationally representative norming sample, or with certain tests or measures, such as the MMPI, particular clinically diagnostic samples.2The brief overview presented here draws on the works of De Ayala (2009) and DeMars (2010), to which the reader is directed for additional information.

The Kaufman Assessment Battery for Children (KABC) is a clinical instrument (psychological diagnostic test) for assessing cognitive development. Its construction incorporates several recent developments in both psychological theory and statistical methodology. The test was developed by Alan S. Kaufman and Nadeen L. Kaufman in 1983 and revised in 2004. The KABC also gives ...