GERIATRIC PSYCHOLOGICAL TESTING DECISION SUPPORT GUIDE

Introduction

Testing of personality characteristics, intelligence, aptitude, or achievement is sometimes performed to assist with diagnosis and/or treatment planning for behavioral health disorders. However, the clinical utility and cost-effectiveness of formal psychological testing have been questioned. In most situations, clinical interviews and brief assessment instruments provide sufficient information for diagnosing behavioral health disorders and determining the most appropriate treatment. Occasionally, a particular psychological test may be helpful for addressing a very specific clinical question; these situations must be reviewed on a case-by-case basis.

Because all potentially appropriate psychological tests cannot be listed in this document and because a partial list might be misinterpreted as a compendium of approved tests, no such list is provided. The specific tests cited are provided only as examples, and were chosen based on their popularity. When psychological testing is requested, the secondary reviewer must evaluate the psychometric properties of the particular test requested and its applicability to the specific clinical situation.

Definition

The InterQual® Behavioral Health Level of Care (BH-LOC) Criteria define psychological testing as the administration, scoring, and interpretation of standardized projective, objective personality, intelligence, aptitude, or achievement tests for the diagnosis and treatment of behavioral health disorders when the time spent on test administration, scoring, and interpretation would be authorized or billed separately from diagnostic interviews or psychotherapy sessions. Examples of psychological tests for which administration, scoring, and interpretation are often authorized or billed separately from clinical interviews include the Millon Clinical Multiaxial Inventory-III (MCMI-III), Minnesota Multiphasic Personality Inventory-2 (MMPI-2), Myers-Briggs Type Indicator (MBTI), NEO Personality Inventory-Revised (NEO-PI-R), Personality Assessment Inventory (PAI), Rorschach Inkblot Method, Strong Interest Inventory, Thematic Apperception Test (TAT), and Wechsler Adult Intelligence Scale-IV. The InterQual BH-LOC Criteria definition of psychological testing does not include the use of brief screening or diagnostic tools that generally require fewer than 20 minutes of clinician time to administer, score, and interpret. Such instruments, which are usually administered as part of the clinical interview or are self-administered outside of the session, should not be reviewed separately as psychological testing.[1] Rarely, a provider may request a battery of brief instruments (the administration, scoring, and interpretation of which will require more than 20 minutes total time) and ask that this combination be authorized as psychological testing; such requests should be considered on a case-by-case basis.

This paper does not address neuropsychological testing for medical conditions such as dementia, stroke, substance-related brain damage, traumatic brain injury, epilepsy, or brain tumor. For information on this topic, see the Neuropsychological Testing subset in the InterQual BH-LOC Geriatric Psychiatry Criteria.

Overview

Any proposed psychological testing should address a specific clinical question.[2-5] Formal psychological testing is not clinically indicated for routine screening or assessment of behavioral health disorders.[2, 6] When a provider requests authorization for psychological testing, he/she may cite purposes that are imprecise and/or clinically questionable. Examples of stated purposes that compel further exploration include:

- Determining a diagnosis or differential diagnosis
- Establishing or ruling out comorbid psychiatric conditions
- Clarifying symptoms and/or functional impairment
- Treatment selection or individualized treatment planning
- Delineating personality constructs such as dependency or narcissism
- Forensic applications (e.g., competency, pre-sentencing, or child custody evaluations)
- Academic or occupational applications (e.g., required for graduate school admission or hiring into a public safety position).

In such cases, the clinician should be asked for more specific information about the focus or expected benefits of testing.

Psychometric Properties

When evaluating a request for psychological testing, the empirical support for the proposed test(s) and its psychometric properties must be ascertained, particularly as they relate to the patient (i.e., norms) and the specific clinical question to be answered (i.e., validity). To be considered for use, a test must first demonstrate reliability.[6, 7] Inter-rater reliability, internal consistency, and test-retest reliability are three common methods of demonstrating the consistency of test results.[7, 8] A test must also be properly normed—that is, the population upon which the interpretation of scores is based must, in general, represent the population for which the test is used.[7, 9, 10] For example, if the interpretation of test scores is based on a sample of European- and African-American persons, one cannot presume that the test is also appropriate for Americans of Hispanic or Asian descent. Likewise, a test normed on a population under age 65 may not be appropriate for geriatric patients.
Test validity is much more difficult to demonstrate than test reliability, as many psychological tests do not have a single specific purpose, and any chosen purpose is difficult to validate.[7, 8, 11] Whereas a medical test designed to identify cancer can be validated based on pathology results, the validity of measures of psychological constructs is not so easily confirmed. Although one frequently-cited article on this topic contends that psychological testing is as valid as medical testing, serious problems with the authors’ methodology and interpretation of results obviate that conclusion.[3, 12-14]

In the context of psychological testing, the question, "Is this test valid?" is inadequate. Instead, one must ask, "Is this test valid for the specific question I intend it to answer?" For example, if one proposes to use the MMPI-2 to screen for personality disorders, one must ask if any research studies have demonstrated that the MMPI-2 can accurately do so. One must additionally ask what criteria were used in those studies to determine whether the MMPI-2 results were accurate, as the reliability and validity of interview-based diagnoses (which are often used as criteria in these types of studies) can also be problematic.[8, 15]

The test’s positive and negative predictive powers are also key considerations when evaluating test validity. If the purpose of a test is to screen for or to rule out certain conditions, particularly those that carry a high risk if they are overlooked (e.g., suicidality), the test should have high negative predictive power. That is, its false-negative rate should be exceptionally low, even at the expense of increased false-positive findings. If it is more important to confirm certain conditions (e.g., to be absolutely sure of a diagnosis before commencing aggressive intervention), then the test should have high positive predictive power and generate relatively few false-positive findings.

Because of the difficulty of performing validity studies and to the almost unlimited number of clinical reasons for which psychological testing is performed, projective and objective personality tests are frequently used to address clinical questions for which they have never been adequately validated. Some clinicians cite personal experience as justification for administering a psychological test with no proven validity for the question at hand.[16] However, this is inconsistent with evidence-based practice and with research studies that have demonstrated little or no association between amount of clinical experience and the accuracy of inferences drawn from psychological tests.[17, 18]

Incremental validity, clinical utility, and cost-effectiveness are significant factors in determining the necessity of psychological testing, yet little research has been done on these topics for even the most popular instruments.[7, 8, 19-21] Regardless of whether a requested test has demonstrated reliability, relevant norms, and validity for the stated clinical question, the issue remains as to whether the test contributes additional information that cannot be uncovered by the clinical interview or by more efficient assessment techniques.[20] Moreover, if the test does contribute additional information, one must assess the value of this information. Incremental validity can best be defined by the following question: how much more valid does the clinician’s judgment become with the addition of psychological test results? If the provider can establish the diagnosis and devise a treatment plan based on the findings of clinical interviews and rating scales, will the results of formal psychological testing contribute anything of importance? Although it can be argued that the valid use of a reliable test almost always provides some additional information, the benefit of that information must be weighed against the cost of the testing.[22] In many cases, there is insufficient evidence that formal psychological testing provides enough added value to justify its cost.[19, 20, 23] In fact, if the test is not valid for the stated clinical question, it may even have negative incremental validity in the hands of clinicians—that is, the test results could make the clinician’s judgment less valid.[7, 19, 20]

**Assessment Methods**

For adults of all ages, the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) has demonstrated excellent validity for determining primary and comorbid diagnoses, and should be considered the most appropriate tool for this purpose.[15, 19, 23, 24] Collaborative information obtained from family members, employers, or significant others can enhance the accuracy of the diagnosis.[3, 7] Medical testing, including thyroid tests or toxicology screening for drugs of abuse, may also be indicated based on the findings of clinical interviews and rating scales, will the results of formal psychological testing contribute anything of importance? Although it can be argued that the valid use of a reliable test almost always provides some additional information, the benefit of that information must be weighed against the cost of the testing.[22] In many cases, there is insufficient evidence that formal psychological testing provides enough added value to justify its cost.[19, 20, 23] In fact, if the test is not valid for the stated clinical question, it may even have negative incremental validity in the hands of clinicians—that is, the test results could make the clinician’s judgment less valid.[7, 19, 20]

Occasionally, formal psychological testing may be useful in the evaluation of exceptionally complex clinical presentations. In these cases, the choice of test must reflect the specific clinical question(s) raised by the clinical interview and other preliminary assessment procedures. Psychological testing may be particularly appropriate when patient self-report cannot be trusted, such as when malingering is suspected and collaborative information is indeterminate. However, the empirical support and psychometric properties of the proposed test must be reviewed, as only certain instruments have demonstrated validity for this situation and even these may be prone to a high false-positive rate.[7, 29-32] Examples of tests that have demonstrated effectiveness and efficiency in cases of suspected malingering are the MMPI-2 and, when the patient is suspected of feigning attention-deficit/hyperactivity disorder, the Continuous Performance Test.[33-45]

Once an initial diagnostic impression is formed, rating scales and inventories can be useful for quantifying symptoms and assessing the degree of impairment.[2, 7, 25, 46-92] These instruments typically require less than 20 minutes of clinician time to administer, score, and interpret, and should not be reviewed as psychological testing.[1] Examples include:

- Global measures such as the Symptom Checklist-90-Revised (SCL-90-R) and Brief Psychiatric Rating Scale (BPRS)
- Anxiety inventories or scales (e.g., Beck, Hamilton, State-Trait)
- Depression, hopelessness, and suicidality scales (e.g., Beck, Hamilton, Reynolds, Geriatric, PHQ-9)
- Substance use assessments such as the Alcohol Use Disorders Identification Test and the Michigan Alcohol Screening Test.
- Neurological screening tools such as the Bender Visual-Motor Gestalt Test II, Trail Making Test, and Quick Neurological Screening Test-II
Because these instruments are targeted and efficient, and most are empirically supported and psychometrically sound, formal psychological testing (as previously defined) is usually not necessary to elucidate symptoms or functional impairments. If personality, projective, intelligence, aptitude, or achievement testing is proposed as an adjunct to structured interviews and brief instruments, the incremental validity of the additional information provided by these tests must be weighed against the time and cost of testing.[4, 5, 7, 20, 93] In most situations, the additional expense of formal psychological testing is unlikely to be justified.[23]

When the clinical formulation is complete, evidence on the effectiveness of each potential treatment alternative (in the context of the presenting psychopathology) should form the basis for treatment planning. When multiple evidence-based treatments are available, patient-specific factors must be considered. However, in the absence of diagnostic uncertainty, there is no research demonstrating that the use of projective, objective personality, intelligence, aptitude, or achievement tests provides information useful for treatment selection beyond what can be obtained through interviews or brief rating scales.

In some cases, psychological testing may be proposed to evaluate psychological constructs such as ego strength, dependency, narcissism, and antisocial tendencies. However, there exists no body of research demonstrating that defining and measuring such characteristics improves treatment outcome.[7] When testing is requested for this purpose, the proposed action to be taken (based on the test results) should be closely evaluated so that the practical value of the proposed testing can be assessed.

Once treatment is underway, clinicians should periodically re-evaluate the patient’s condition. Rating scales are effective means of measuring patient progress, and are the predominant means of outcome measurement in clinical research.[46, 94-98] However, formal psychological testing may be appropriate if (a) the patient is not making progress, (b) questions remain about the primary or potentially co-occurring diagnoses, and (c) less-intensive assessment methods have failed to clarify the situation. If formal psychological testing is planned to assess routine progress or post-therapy outcomes, further investigation is recommended. Although testing in these situations may be appropriate for research or programmatic needs (e.g., quality initiatives, treatment outcome studies), it is usually not clinically necessary for the patient.

Certain tests have proven validity for predicting occupational and graduate school performance, and psychological testing is frequently used in court-ordered evaluations.[7, 23] However, in the context of a health plan, the occupational or forensic use of psychological testing should be evaluated thoroughly to determine if such testing is clinically necessary to diagnose or treat a psychiatric or substance use disorder.

Specific Tests

Objective personality tests tend to have a stronger empirical base and to be more psychometrically sound and efficient than projective instruments. The MMPI-2 offers a large number of scales validated for a broad variety of clinical questions.[99-101] Although these questions can usually be more efficiently addressed by other assessment methods, the MMPI-2 is occasionally helpful when symptoms are complex and the diagnosis is unclear.[102] The recently developed Restructured Clinical (RC) scales for the MMPI-2 may possess superior convergent and discriminant validity compared with the standard MMPI-2 scales, but the evidence thus far is preliminary and requires further investigation.[40, 99-101, 103] The MCMI-III and PAI have some validity for the identification of certain personality disorders.[104-114] These tests, like the MMPI-2, raise cost-benefit issues; but in the case of an equivocal presentation of a possible personality disorder, when other assessment techniques have proven inadequate and the identification of a personality disorder is important to treatment, the MCMI-III or PAI may be appropriate.

Projective tests include the TAT, the Rorschach Inkblot Method, sentence completion tests, and various drawing tests. Although clinicians have asserted the usefulness of these measures, the TAT and most sentence completion and drawing tests are not standardized, and the clinical value of the information they do yield (such as personality traits) has not been empirically established.[7, 115]

The Rorschach Inkblot Method is sometimes scored using Exner’s Comprehensive System, which allows for investigation of its reliability, validity, and norms. Despite this, the Rorschach is perhaps the most controversial of psychological tests.[7, 116] Disagreement exists about its reliability, the validity of its scales and indices, and whether the Comprehensive System’s norms are accurate.[7, 117-129] Concerns have been raised that the norms tend to overpathologize respondents (that is, make them appear to be more psychologically ill than they actually are) and that these norms may be especially problematic when applied to ethnic minorities.[23, 121-124, 128-130] The Rorschach is not intended to provide a DSM-IV diagnosis, and there is currently no scientific evidence that use of the Rorschach improves treatment outcomes.[23, 118-124, 131-135] Even if serious concerns about the norms are set aside, scales that do appear to provide relevant information, such as the Schizophrenia Index (SCZI), are often unnecessary because other assessment methods are more efficient at providing such information.[124, 129] As with all psychological tests, the Rorschach’s incremental validity for the question at hand must be evaluated in relation to the cost of the test.[123]

Personality or preference measures designed for normal individuals, such as the California Psychological Inventory, NEO-PI-R, Sixteen Personality Factors, Strong Interest Inventory, and Taylor-Johnson Temperament Analysis, are not valid for diagnosing or treating psychiatric or substance use disorders.[136] In rare instances, they may be employed as a part of a treatment plan that targets relationship issues or life choices as the source of the dysfunction (e.g., marital therapy to treat adjustment disorder, occupational evaluation to treat depression related to job dissatisfaction). All requests for such testing should be closely scrutinized to determine whether it is truly clinically necessary for treating the psychiatric or substance use disorder.
Practical Recommendations

1. When reviewing a request for psychological testing, first clarify the specific question the testing is intended to answer.

2. Next, review the assessment methods that have been used to date. If assessment procedures that appear to be more efficient and/or appropriate have not yet been performed, a discussion with the provider may be warranted. The following are examples of such situations:
   - The patient was referred to the provider by his/her employer or employee assistance program, but the provider has not requested information from the referral source.
   - Indications of substance abuse are present, but the provider has not performed (or has not referred the patient for) a chemical dependency assessment.
   - Symptoms include debilitating fatigue, confusion, severe headaches, or psychomotor agitation, but the patient has not undergone (or has not been referred for) a medical evaluation.
   - The patient has experienced a serious medical condition, but the provider has not attempted to confer with the patient’s physician or to obtain medical records. Acute medical conditions that can complicate the diagnosis and treatment of psychiatric disorders include myocardial infarction, stroke, traumatic injury (including burns), major surgery, and a new diagnosis or recurrence of cancer.
   - The patient has a chronic medical condition, but the provider has not attempted to collaborate with the patient’s physician. Chronic medical conditions that can complicate the diagnosis and treatment of psychiatric disorders include chronic obstructive pulmonary disease, sleep apnea, coronary artery disease, heart failure, diabetes, Crohn’s disease, renal failure, frequent migraine or cluster headaches, chronic low back pain, complex regional pain syndrome, severe osteoarthritis, rheumatoid arthritis, fibromyalgia, multiple chemical sensitivities, and chronic fatigue syndrome.

   If it appears that the case-specific question could be more efficiently or appropriately addressed by assessment procedures other than formal psychological testing, consider recommending that such procedures be performed first. Organizational policy will determine whether and how such a recommendation will be communicated.

3. If more efficient assessment procedures are not indicated or have proven inadequate to answer the specific question, determine the importance of the question for diagnosis and treatment planning. The provider should be able to clearly elucidate how the test results will affect treatment; if he or she cannot do this, the necessity of testing can be legitimately questioned.

4. Verify the reliability of the proposed test, its validity for the specific question, and the appropriateness of its norms for the patient. Such information is usually available in the test publisher’s documentation, and studies and literature reviews on the reliability and validity of the most commonly used tests are published in psychological journals. Whether the provider is asked to submit this information or whether the secondary reviewer must obtain it is left to organizational policy.

5. If the case is particularly difficult, the clinical question is specific and important, the proposed test is known to be valid for that question, and the test is normed appropriately for the patient, then psychological testing may be appropriate. The number of hours authorized for any psychological test should be based on the time required to administer, score, and interpret the instrument. (Note that test publishers may indicate only administration times; in such situations, additional time must be provided for scoring and interpretation.) This document does not include a table that assigns hours for specific tests, because all potentially-appropriate psychological tests cannot possibly be included, and because readers hastily scanning this document could easily misinterpret such a table as a list of approved tests. In addition, reimbursement issues may be affected by whether psychologists, other clinical staff, non-clinical staff, and/or computers are used for administration and/or scoring; and these factors are sometimes subject to state regulations. Moreover, the time needed to interpret certain psychological instruments can vary greatly based on the extent to which a patient’s responses are pathological. For these reasons, organizations are encouraged to develop their own guidelines regarding the number of hours authorized.
References


37. Tsushima WT and Tsushima VG, Comparison of MMPI-2 validity scales among compensation-seeking Caucasian and Asian American medical patients. Assessment, 2009. 16(2):159-164. (III)


118. Garb HN, Call for a moratorium on the use of the Rorschach Inkblot Test in clinical and forensic settings. Assessment, 1999. 6(4):313-318. (V)