CHILD 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 the test’s 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 Childhood Apperception Test (CAT), Kaufman Assessment Battery for Children-Second Edition (KABC-II), Peabody Individual Achievement Test-Revised (PIAT-R), Personality Inventory for Children-Second Edition (PIC-2), Raven’s Progressive Matrices, Roberts Apperception Test for Children-Second Edition (Roberts-2), Rorschach Inkblot Method, Vineland Adaptive Behavior Scales-Second Edition (Vineland-II), Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV), Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III), Wide Range Achievement Test 4 (WRAT4), and Woodcock-Johnson III. 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. 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 traumatic brain injury, anoxic or substance-related brain damage, epilepsy, and brain tumor; or for language disorders or pervasive developmental disorders. For information on this topic, see the Neuropsychological Testing subset in the InterQual BH-LOC Child Psychiatry Criteria.

Overview

Any proposed psychological testing should address a specific clinical question.[1-3] Formal psychological testing is not clinically indicated for routine screening or assessment of behavioral health disorders. 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., child custody evaluations, abuse investigations)
- Academic applications (e.g., developing or revising an individualized education plan)

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.[4, 5] Inter-rater reliability, internal consistency, and test-retest reliability are three common methods of demonstrating the consistency of test results.[4-6] 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.[4, 5, 7, 8] For example, if the interpretation of test scores is based on a sample of European- and African-American children, one cannot presume that the test is also appropriate for those of Hispanic or Asian descent. Likewise, a test normed on a population over age 12 may not be appropriate for younger patients.

Test validity is much more difficult to demonstrate than test reliability, as many psychological tests do not have a single specific
In the context of psychological testing, the question, “Is this test valid?” is insufficient. 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 Personality Inventory for Children-Second Edition (PIC-2) to rule out depression, one must ask if any research studies have demonstrated that the PIC-2 can accurately do so. One must additionally ask what criteria were used in those studies to determine whether the PIC-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.[6, 8, 13]

The test's positive and negative predictive powers are also key considerations when evaluating test validity.[8] 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 performing validity studies and the almost unlimited number of clinical reasons for which testing is performed, psychological tests are frequently used for purposes for which they have never been adequately validated. Some clinicians cite personal experience as justification for administering a psychological test with no proven validity.[14] 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.[15, 16]

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:[4, 8, 17, 18] 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. 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.[8, 17, 19] In many cases, there is insufficient evidence that formal psychological testing provides enough added value to justify its cost.[17, 20] 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.[4]

Assessment Methods

Structured interviews to be used with children and their parents, including the Diagnostic Interview Schedule for Children (DISC-IV), the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS), and the Anxiety Disorder Interview Schedule for Children and Parents (ADIS-IV-C/P), have demonstrated good utility for determining primary and comorbid diagnoses.[8, 20-32] For very young children, the Berkeley Puppet Interview may be useful.[33] Obtaining collaborative information from teachers or others involved in the patient's life is also recommended to validate the patient's self-reported symptoms and to provide a more thorough history.[4, 7, 8, 34-46] Medical testing, including thyroid tests or toxicology screening for drugs of abuse, may be indicated based on the clinical interview.[38-40, 44] Research does not support the use of projective or objective personality testing as a primary method of establishing or ruling out DSM-IV psychiatric diagnoses.[23, 37-40, 43-46] 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.

Once an initial diagnostic impression is formed, rating scales and inventories can be useful for quantifying symptoms and assessing the degree of impairment.[4, 8, 21-24, 41-50] These instruments may be completed by the patient, parents, and/or others (such as teachers), and typically require less than 20 minutes of clinician time to administer, score, and interpret; therefore, they should not be reviewed as psychological testing. Examples include the Childhood Autism Rating Scale-Second Edition (CARS2), Child Behavior Checklist, Children's Depression Inventory, Childhood Depression Rating Scale-Revised, Children's Yale-Brown Obsessive Compulsive Scale, Conners Rating Scales-Revised, DeVereux Scales of Mental Disorders, Hopelessness Scale for Children, Multidimensional Anxiety Scale for Children (MASC), Pediatric Anxiety Rating Scale (PARS), and Screen for Child Anxiety Related Emotional Disorders (SCARED).[22-25, 42-47, 50-82] For very young children, the MacArthur Health and Behavior Questionnaire (HBQ) or the Preschool Feelings Checklist may be useful.[33, 83, 84] The use of two different rating scales may provide more clarity without requiring significant additional clinician time.[50] 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.[2-4, 8, 41] In most situations, the additional expense of formal psychological testing is unlikely to be justified.[20]

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.[7, 20, 85] 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.
Once treatment is underway, clinicians should periodically re-evaluate the patient's condition.[7, 49] Rating scales are effective means of measuring patient progress, and are the predominant means of outcome measurement in clinical research.[7, 8, 23, 24, 37, 43, 46, 49, 50, 65-69, 86, 87] 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.

Psychological testing may also be proposed to address legal matters such as child custody evaluations or assessment of potential child abuse.[4, 88, 89] The forensic use of psychological testing should be evaluated thoroughly to determine if the testing is clinically necessary to diagnose or treat a mental disorder. As noted previously, formal psychological testing is not designed to provide a diagnosis, and is rarely indicated for treatment planning. Projective drawings have been used to investigate whether a child has been abused; however, the reliability and predictive validity of these techniques remain questionable.[20, 89-93]

Specific Tests

Objective personality tests, such as the PIC-2, tend to have a stronger empirical base and to be more psychometrically sound and efficient than projective instruments, which include the Roberts-2, the Rorschach Inkblot Method, sentence completion tests, and various drawing tests.[4, 94] Most sentence completion and drawing tests are not standardized.[4, 95] For children under age 13, the clinical value of the information provided by projective and objective personality testing has not been empirically established.

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. Disagreement exists about its reliability, the validity of its scales and indices, and whether the Comprehensive System’s norms are accurate.[4, 20, 96-102] 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 children and ethnic minorities.[20, 99-103] 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.[12, 20, 100, 101] 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.[102, 104] 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.[12]

Although developmental, aptitude, achievement, or intelligence tests may be useful for identifying or monitoring suspected developmental delays or learning disabilities, federal law requires that local education and early intervention agencies provide testing for minors who may have disabilities or special educational needs (see the Individuals with Disabilities Education Improvement Act of 2004).[105-107] Private testing may be requested when schools or early intervention agencies lack the resources to provide timely, sufficient, or appropriate testing. Although health plans often recognize the limitations of school system resources and are willing to authorize psychoeducational testing in these situations, development of a specific organizational policy is recommended to address local alternatives and reimbursement issues.

Neuropsychological testing is sometimes proposed for the diagnosis of attention-deficit/hyperactivity disorder. However, such testing is not routinely indicated or necessary.[44, 45, 108] Continuous performance tests, including computerized instruments such as the Test of Variables of Attention (TOVA), have demonstrated inadequate specificity and sensitivity in children and adolescents.[45, 55, 109-111]

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 by school personnel, but the provider has not yet obtained information from the referral source.
   - The patient has experienced a serious medical condition, such as a head injury or cancer, but the provider has not attempted to confer with patient’s physician or obtain medical records.

   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.


100. Wood JM, Lilienfeld SO, Nezworski MT, et al., Coming to grips with negative evidence for the Comprehensive System for the Rorschach: a comment on Gacono, Loving, and Bodholdt; Ganellen; and Bornstein. J Pers Assess, 2001. 77(1):48-70. (V)
