3 Selection of Psychological Measures and Associated Administration, Scoring, and Reporting Technology for Use in Pediatric Primary Care Settings

Mark E. Maruish

Test publishers and other developers of psychological measures (e.g., researchers, universities, government agencies) regularly release new psychological test instrumentation that can facilitate and evaluate behavioral health care treatment in primary care settings. Thus, availability of instrumentation for these purposes is not an issue. However, selection of the appropriate instrument(s) for screening, treatment monitoring, or outcomes assessment is a matter requiring careful consideration. Inattention to instrument’s intended use, its demonstrated psychometric characteristics, its limitations, and other aspects related to its practical application can result in a misguided diagnosis and/or treatment and potentially harmful consequences for a patient.

Similarly, various modes and methods of automated administration, scoring, and reporting/interpretation of the results of psychological measures that capitalize on the capabilities of today’s technology also are available. These methods require the same level of consideration and scrutiny as the measures they deliver in order to determine which is the most suitable for the settings in which they are implemented.

The purpose of this chapter is to identify and discuss important factors that should be carefully considered when one sets out to determine which among the available psychological test options is optimal for one or more clinical purposes within a pediatric primary care setting. The chapter is also intended to similarly identify and discuss factors that should be considered in the selection of psychological testing technologies. Considerations for the selection of psychological measures will be addressed first, followed by a discussion of some of the more common technologies that are available for the administration, scoring, and/or reporting of those measures.

Selection of Psychological Test Instruments

The selection of a particular psychological test instrument for use in any setting—including pediatric primary care settings—is a process that should involve careful thought and deliberation on the part of the individual(s) responsible for making the choice. Perhaps the best place to begin the process is asking whether use of a particular psychological test for the intended purpose (e.g., screening a child for depression, monitoring patient recovery over time) is the best solution for the task. While recognizing that psychological tests’ efficiency and objectivity are significant advantages over other methods of clinical information gathering, Urbina recommended doing a cost-benefit analysis similar to one suggested by Goldman (cited in Urbina, 2004), asking questions pertaining to (a) what kind of information will be gained, (b) how that information will be used, (c) how much of that information is available from other sources, (d) what other ways that information can be gained, (e) what are the advantages of using psychological tests in addition to or instead of other information sources, and (f) what are the disadvantages of using tests (e.g., cost, time involved) in addition to or instead of other information sources.
Assuming that the benefits and usefulness of using a psychological test outweigh the use of other information-gathering methods, it is important to remember that no instrument is best for all settings (Bufka, Crawford, & Levitt, 2002; Ogles, Lambert, & Fields, 2002), regardless of its intended purpose. Consequently, there are many factors to consider as part of the process for evaluating the suitability of one or more tests for a particular purpose. Some of these factors are general in terms of their applicability to various types of tests for various purposes; others are specific to instruments that are used for specific purposes and/or with specific patient populations. Both types of factors are discussed in this section.

**General Considerations for Instrument Selection**

Regardless of the type of instrument one might consider using in a pediatric primary care setting, psychologists frequently must choose between many product offerings. But what are the general considerations for the selection of any instrument for psychological testing? What should guide the clinician’s selection of an instrument for a specific purpose? As part of their training, psychologists and other appropriately trained professionals from related specialties have been educated about the psychometric properties that are important to consider when determining the appropriateness of an instrument for its intended use. However, psychometric integrity is just one of several considerations that should be taken into account in evaluating a specific instrument for a specific clinical application.

Finally, one other point is worth noting. In discussing selection criteria of structured or semistructured interviews, Summerfeldt and Antony (2002) noted that “no one instrument best fits the requirements of all clinicians and researchers. When selecting an interview, healthcare workers must consider their specific needs, priorities, and resources” (p. 4). This same caveat holds true not only with regard to interviews but also with regard to the other psychological instruments.

Following are considerations that are recommended for the selection of tests and other assessment instruments for use within pediatric primary care settings, regardless of their intended purpose(s). Some of these may seem obvious, but one would be surprised how easily some of these considerations can be overlooked. Following the general considerations are additional considerations for instruments that will be used for specific purposes or under specific circumstances.

**Relevancy to the Intended Purpose of Assessment**

Evaluating a given psychological test instrument should begin with identifying what is to be measured (Fischer & Corcoran, 1994) and then determining the relevancy of the instrument to the reason for testing the patient. Groth-Marnat (2009) indicated that when selecting a test, the most important consideration is how useful the test is in answering the referral question. This recommendation can be extended to cover other reasons for testing, such as screening for the presence of psychological symptomatology, determining personality characteristics important to facilitating a course of psychotherapy, or for routine treatment monitoring and outcomes assessment. Administering a Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001)—a depression screener—to screen for levels of general distress is not the most appropriate way to accomplish the task unless, of course, this screener has been empirically shown to be effective when used this way. Regardless of its psychometric integrity, a measure lacking empirical support for the intended use should not warrant any further consideration.

**Source of Information**

With the majority of pediatric patients, the clinician will probably want the patient to serve as the source of information (i.e., the respondent to the instrument). Usually, however, a parent or
other adult must also serve as an information source. This is particularly the case with children, but there also are adult cases in which the clinician will have to look to another adult to obtain the desired information (e.g., severely impaired adults, reluctant participants in the assessment process). In those cases, only instruments specifically developed to obtain information about a patient from another person such as a family member (e.g., parent) or other knowledgeable party (e.g., guardian, child’s schoolteacher) should be considered. In these cases, the instrument should be evaluated in the same manner as a patient self-report instrument using the considerations noted in this section. In no instance should the use of a patient self-report instrument be evaluated for use with a collateral respondent based on data (e.g., validity, reliability, norms) obtained directly from actual patient samples.

**Instrument Content**

The content of the test—what the test is asking the respondent to report on—is also important in achieving acceptability to the respondent (e.g., see Bufka et al., 2002), particularly in the cases of adult respondents. Asking questions that are not considered relevant by the parent or, in some cases, the child (i.e., face invalidity) or that are informative but potentially offensive should be avoided. A respondent’s perceptions of what is appropriate or useful to inquire about can have a direct bearing on the probability that they will give a valid response to any or all of the questions being asked.

A related content issue is whether the measure is theory-based or is atheoretical. In some cases, the user may want to use a measure that is consistent with a particular theory he or she espouses or one that is not meant to reflect any particular theory.

**Psychometric Integrity**

Having good psychometric properties is a crucial factor for any assessment instrument (Groth-Marnat, 2009). Thus, in selecting a test, one must ask, “Is the instrument a valid and reliable measure of what it purports to measure?” Also, “Should specific aspects of validity and reliability be attended to more than others when evaluating psychological measures?” There is no one clear answer to this latter question. All relevant aspects of validity and reliability are important to consider in selecting a psychometric instrument, but depending on the intended use, particular types of validity and/or reliability might be given more weight in the decision. For example, if the intended use of an instrument is for treatment monitoring, one would like to see empirical data demonstrating good test-retest reliability. When used for screening, one should closely consider the instrument’s criterion validity and accompanying diagnostic efficiency statistics (discussed later in this section).

Regardless of the intended use of the measure, there are several aspects of an instrument’s reliability and validity that should be evaluated. The instrument’s internal consistency and test-retest reliabilities are important reliability considerations to ensure consistency of item content and stability of test results over time. Important validity considerations include the instrument’s construct validity, which can be evaluated through means such as examination of the instrument’s content, factor analytic studies, correlations with other measures of the same and dissimilar constructs, ability to discriminate groups of individuals with and without a particular trait or condition, and ability to predict a behavior or outcome in the future.

A detailed discussion of desired psychometric properties of psychological measures is beyond the scope of this chapter. The reader is referred to the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) for further guidance on these matters. Others such as Smith and Archer (2014), Urbina (2004),
Cicchetti (1994), and Fischer and Corcoran (1994) provide further discussion and guidance related to the evaluation the validity and reliability of psychological measures being considered for clinical use.

**Availability of Relevant Normative Data**

Tied to the relevancy and psychometric soundness of the instrument is the issue of whether normative data are available for the particular population(s) with whom the measure will be used. Using the Geriatric Depression Scale to screen geriatric patients for depression is quite appropriate. However, use of this same instrument with pediatric populations should be avoided. As Fischer and Corcoran (1994) note, the patient being administered a test should be considered a member of the population on which the test’s norms are based. Moreover, they recommend determining whether the norms are based on a sample that is current or up to date and of a sufficiently large size.

**Comprehensibility of Results**

Another often-overlooked factor in instrument selection is how easily the test results can be understood by the primary care team or other relevant individuals, such as the patient and their family. Aside from the fact that authorization for appropriate treatment may hinge upon the primary care team fully comprehending the implications of the test findings, the psychologist should be able to easily explain the results to the patient and/or his family in order that they more clearly understand the patient’s problems and the implications they have for his or her day-to-day functioning. In some cases, the results may reveal areas of strength, which, if understood by the patient, may facilitate efforts toward positive change.

**Actionable Information**

*Actionable information* generally refers to information that one can do something with, that is, information that gives the care provider clear direction about how to improve the quality of services offered to patients. During the ongoing monitoring of an individual’s progress in treatment, information that is actionable can be used to indicate when treatment isn’t working as well as to suggest the changes that need to be made. However, the patients who provide posttreatment outcomes data that yield actionable information frequently do not benefit from that information, since they are no longer in treatment. Instead, the information that they provide can lead to changes that future patients with similar problems will benefit from. For example, aggregated outcomes data from a multidimensional symptom inventory may show that a clinician does quite well in treating depression, but their efforts yield minimal results with anxiety disorders. As a result, the clinician may decide to obtain more supervised training in the treatment of anxiety disorders or, alternatively, limit his practice to patients with depressive disorders.

**Usefulness across the Continuum of Care**

The ability to use an instrument across different levels of care (LOCs) may be an important consideration in some cases, depending on the range of services the primary care clinician offers and other types of service settings in which the patient is seen (e.g., inpatient). In particular, some pediatric patients with chronic behavioral health problems may be treated at more than one level of health care setting due to a progression of the problem or the waxing and waning of their illness over time. Clinicians wishing to monitor these types of patients over time will want the instrument they select for this task to be capable of assessing them on the variables of interest.
(e.g., symptomatology, social functioning), regardless of their fluctuating mental and behavioral status, the type of treatment being used, or the setting in which they are being treated.

**Ease of Use**

One selection consideration that many professionals may overlook, or at least minimize in their importance, is the ease with which the test is administered and scored by themselves or (in cases in which the psychologist is not going to be engaged in these activities) other staff members, such as physicians, nurses or administrative support staff. Considerations here can include the psychologist’s training in the use of the instrument (Groth-Marnat, 2009) or how easily the psychologist or other person(s) who will be administering, scoring, and/or interpreting the instrument can become skilled in its use (see Bufka et al., 2002; Fischer & Corcoran, 1994; Newman, Rugh, & Ciarlo, 2004).

**Clinical Utility**

There are several ways in which one can define and assess the utility of a psychological measure in clinical settings. For instance, Fischer and Corcoran (1994) indicate that determining the utility of an instrument requires that one consider both the practical advantage of the information the instrument yields and whether that same information could be obtained in another way. Smith and Archer (2014) define incremental validity as the primary form of clinical utility, which they define as “the ability of a measure to add a new form of information or improve classification accuracy over and above another established measure of the same construct” (p. 24). In addition to its ease of use and cost effectiveness (i.e., the information the measure yields compared to other measures), Smith and Archer identify the measure’s diagnostic efficiency as a means of evaluating its clinical utility. Diagnostic efficiency has to do with how well an instrument can correctly classify those with the disorder of interest from those without the disorder, typically by comparing the patient’s obtained test score to an established classification cutoff score.

Commonly calculated diagnostic efficiency statistics include sensitivity, specificity, positive predictive power, negative predictive power, overall predictive power, and kappa. Note that the base rate of a particular disorder in a given setting will affect the instrument’s predictive power for that setting. And as Urbina (2004) points out, in situations in which the base rate for a specific disorder is extremely low, the cost of using the test for identifying those with that disorder might not be justified by the relatively small gains that can be achieved by its use. Moreover, Urbina cautions that one must recognize that in actual clinical settings, classifications and other clinical decisions are not made solely on the basis of test results. The knowledge and judgment of the clinician also come into play.

**Brevity**

One of the most important characteristics of any instrument being considered for use in pediatric primary care settings is that it is brief. Lengthy instruments do not fit well in a primary health care delivery system in which time-limited, problem-oriented intervention is the primary approach to patient treatment. In addition, to maximize the patient’s cooperation with the testing process, the length of the instrument should be acceptable to most patients. This is particularly important with pediatric populations. Keep in mind that what psychologists consider “short” for a test may seem unreasonably long to the patient. The use of lengthy, expensive instruments or batteries of instruments that either provide little useful information or represent overkill with a patient being seen for a limited number of sessions is of little value in primary care settings. An exception may be found in outcomes assessment systems. Generally, however, the shorter the instrument the better.
In addition to the length of the measure or the time for the patient to complete it, there are two other aspects of brevity: the time to score it and the time to review the information and interpret the results (Bufka et al., 2002). These may become important matters to consider, depending on the test and whether automated (i.e., computer-based) scoring and/or interpretive services are available to the clinician. Ideally, one would want to administer, score, and interpret the results of the test before the patient leaves the office.

It is important to note that brevity is meaningless unless the instrument is valid and reliable. In fact, the briefer the instrument, the more one should be concerned with its psychometric properties given that in general, brief instruments tend to be less reliable and valid than longer instruments. Thus, although the psychometric characteristics of all psychological instruments used in pediatric primary care and other settings should be carefully evaluated, particular scrutiny is called for with brief measures of any construct.

**Reading Level**

For instruments that are completed via a paper and pencil, online, or any other electronic format in which items are administered via visual/written presentation, reading level is a major consideration. In general, developers of tests and other assessment instruments have become more sensitive to the issue of reading level vis-à-vis the intended patient population’s expected reading level. Part of this may be due to the fact that software is available to easily determine the reading level of material, thus providing a means of identifying problematic text. Despite these efforts, one is still amazed at the reading level of some instrumentation that is intended for general pediatric and adult patient audiences.

Pediatric primary care settings provide care to patients and parents with various levels of reading ability. Thus, whether the respondent be the child/adolescent or the adult, the recommendation is to select an instrument with a minimum required reading level that is commensurate with the reading level of the respondent. Citing Manly, Jacobs, Touradji, Small, and Stern, McHugh and Behar (2009) note that level of educational attainment (i.e., grades completed) cannot serve as a valid estimate of reading grade level. For parents or other adults who are participating in the assessment of the pediatric patient, they recommend that an instrument such as the Rapid Estimate of Adult Literacy in Medicine (Davis et al., 1993) be used to make such a determination. For the pediatric patient, one can turn to reading comprehension subtests from commercially available instruments such as the Wide Range Achievement Test, Fifth Edition (WRAT-5; Wilkinson & Robertson, 2017) or the Kaufman Test of Educational Achievement, Third Edition (KTEA-3; Kaufman & Kaufman, 2014) as means for estimating the child or adolescent’s reading level. For patients reading below the instrument’s recommended minimal reading level, alternative solutions may include appropriate instruments that have available an oral administration mode, such as by audiotape or an interactive voice response (IVR) system (discussed in what follows).

**Cost**

Cost is always a factor. When funds for psychological testing are limited, one does not want to spend a great deal of that money on expensive testing forms or scoring services. Fortunately, valid, reliable, and useful instruments are becoming more available for little or no cost. Some of these are instruments that have been in the public domain for a number of years; other cost-free instruments have become available only during the past few years. Others may be provided by their developers for a nominal licensing fee. Note that Groth-Marnat (2009) suggests that use of computer-administered measures may help to lower the cost of some instruments.
Overall Feasibility and Practicality

Slade, Thornicroft, and Glover (1999) identified feasibility as a consideration in test selection. In this discussion’s context, they defined feasibility as “the extent to which [the instrument] is suitable for use on a routine, sustainable and meaningful basis in typical clinical settings, when used in a specified manner and for a specified purpose” (p. 245). Feasibility, as conceptualized here, actually is comprised of six characteristics, five of which—brevity, availability, relevance, simplicity, acceptability—are generally comparable to criteria identified earlier by others. What is unique in this conceptualization is the sixth criterion, value. Value is said to occur when the benefits of using the instrument exceed the costs associated with learning how to use the instrument, implementing its use, and analyzing, presenting, and interpreting the resulting data. Although Slade et al. discussed feasibility as it relates to being a desirable criterion for outcomes instrumentation, the characteristics they identified as being part of this construct are in fact desirable characteristics for psychological assessment instruments regardless of how they are used (e.g., screening, treatment monitoring) in a clinical setting.

Related to an instrument’s feasibility is its practicality. Fischer and Corcoran (1994) identified several aspects to consider when judging an instrument’s practicality. In addition to some of the already-discussed considerations (i.e., brevity, content that is acceptable and understandable to the patient, utility, ease of scoring) they point out the importance of the instrument as being sensitive to change but nonreactive. Reactivity here implies that the act of measurement itself causes changes in the patient. In addition, Fischer and Corcoran recommend using instruments that are relatively direct, that is, those whose scores are signs of problems, not symbols of problems. For example, instruments that ask about the experience, intensity, or frequency of a problem are considered to be direct.

Translated and Adapted Versions of Instruments

Some patients seen in pediatric primary care and other health care settings present challenges in terms of selecting an appropriate measure for administration. One such group is the non–English-speaking population of pediatric patients and their parents/guardians and those with English as their second language. Health care settings have already begun to establish means of overcoming the language barrier that can prevent the provision of quality medical care. Bilingual staff members, translators at the ready, and materials printed in commonly encountered languages (e.g., Spanish, Chinese, Arabic) provide relatively simple and effective solutions to the problem. However, one area in which finding an easy solution can be problematic is that involving the use of psychological tests. Simply translating tests that are commonly used in the psychologist’s practice setting or finding an existing translated version of the test (e.g., by word of mouth or on the Internet) without a clear understanding of its psychometric properties or how it was developed is not an acceptable option.

Procedures for and issues related to the development of translated or otherwise adapted versions of existing psychological tests have been presented and discussed by several assessment experts and organizations. In particular, the reader is directed to the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) as well as to the guidelines of the International Test Commission (2001, 2005; also see Acevedo-Polakovich et al., 2007) and the work of Geisinger (1994; Geisinger & McCormick, 2013). Another source of information about translation procedures is Bullinger et al. (1998), who describe the multistage methodology used by the International Quality of Life Assessment (IQOLA) Project. The IQOLA Project sought to develop 14 translations or adaptations of the SF-36 Health Survey that were conceptually equivalent to the U.S. source instrument as well as culturally and linguistically relevant (see Maruish, 2011). Thus, evaluation of any translated or adapted version of a psychological instrument for
clinical use should at least include an assessment of how the development of the translation or adaptation compares to the procedures espoused or utilized by one or more of the cited sources.

For those requiring help in determining the appropriateness of a particular test translation or adaptation, Fernandez, Boccaccini, and Noland (2007) offer a four-step approach to identifying and selecting translated tests. Briefly, Step 1 is identifying translated tests offered by test publishers. Step 2 involves identifying research that involves the use of translated tests, and then confirming that the research applies to the patient (Step 3). In Step 4, one must determine the level at which the research supports the use of the test with the patient(s) in question through available equivalency (construct, metric, and functional) research. Acknowledging the lack of empirical support for many translated tests at the time, Fernandez et al. also provided recommendations for how to deal with such limitations in realistic testing situations. Moreover, the authors identify cultural issues beyond language that should be addressed during the process of test selection.

The general considerations just discussed are summarized in Table 3.1.

<table>
<thead>
<tr>
<th>Instrument Feature</th>
<th>Important Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevancy to the intended purpose of the assessment</td>
<td>• Is appropriate for measuring the targeted domain(s) in the targeted population(s)</td>
</tr>
<tr>
<td>Source of information</td>
<td>• Was developed to obtain information from the desired source of information (e.g., obtaining patient information from a parent using an instrument developed for administration to parents)</td>
</tr>
<tr>
<td>Instrument content</td>
<td>• Has no irrelevant, inappropriate, or offensive item content</td>
</tr>
<tr>
<td>Psychometric integrity</td>
<td>• Meets generally accepted standards for validity and reliability</td>
</tr>
<tr>
<td></td>
<td>• Has demonstrated responsiveness (for individual data) and/or sensitivity (for group data) to changes in patient status</td>
</tr>
<tr>
<td>Availability of relevant normative data</td>
<td>• Has norms that are appropriate for the targeted population</td>
</tr>
<tr>
<td>Comprehensibility of results</td>
<td>• Results can be easily understood by the provider, patient, family members, and other relevant stakeholders</td>
</tr>
<tr>
<td>Actionable information</td>
<td>• Provides the clinician with information that can be used to improve services to the patient</td>
</tr>
<tr>
<td>Assessment across the continuum of care</td>
<td>• Is appropriate for use with patients receiving care at any level of service (e.g., inpatient, outpatient, partial hospitalization)</td>
</tr>
<tr>
<td>Ease of use</td>
<td>• Is easy to administer, score, interpret, and provide feedback</td>
</tr>
<tr>
<td>Clinical utility</td>
<td>• Provides information that cannot be obtained any other way</td>
</tr>
<tr>
<td></td>
<td>• Yields classification accuracy that is better than other measures of the same construct</td>
</tr>
<tr>
<td></td>
<td>• Is cost-effective and easy to use</td>
</tr>
<tr>
<td>Brevity</td>
<td>• Is considered short from the patient’s perspective</td>
</tr>
<tr>
<td>Reading level</td>
<td>• Requires no higher than an eighth-grade reading level, with sixth grade or lower preferable</td>
</tr>
<tr>
<td></td>
<td>• Or, can be administered via another mode that does not require reading (e.g., audio tape, IVR) and yields comparable results</td>
</tr>
<tr>
<td>Cost</td>
<td>• Inexpensive to use for multiple administrations to a single patient</td>
</tr>
<tr>
<td>Overall practicality and feasibility</td>
<td>• Given all considerations, is practical for use in the intended setting, with the intended population, for the intended purpose(s)</td>
</tr>
<tr>
<td>Translated and adapted versions</td>
<td>• Was developed according to commonly accepted procedures and guidelines</td>
</tr>
<tr>
<td></td>
<td>• Has met psychometric and other standards for equivalency with the original version of the instrument</td>
</tr>
</tbody>
</table>

Adapted from Maruish (2013) with permission.
Application-Specific Considerations for Instrument Selection

In addition to the general considerations, the selection of test instruments for specific purposes will sometimes require other considerations.

Instrumentation for Screening and Diagnosis

Screening for the likelihood of the presence of disorders or (more generally) for the need for additional assessment requires considerations that do not necessarily apply to instruments when they are used for the other assessment purposes (e.g., treatment planning, personality description). When selecting tests for use in screening for the presence of general psychopathology or specific disorders in pediatric primary care settings, there are a number of considerations related to sensitivity and specificity that must play into the decision. In general, Ficken’s (1995) requirements of instruments used for screening were noted to include (a) high levels of sensitivity and specificity to DSM or ICD diagnostic criteria; (b) an administration time of no more than 10 minutes; and (c) an administration protocol that easily integrates into the organization’s workflow. Much like Derogatis’s (2017) sequential screening process, cases testing “positive” on the screener then would be administered one or more “second-tier” instrument(s) to establish severity and a specific diagnosis.

Narayana and Wong (2015) also discussed the importance of sensitivity and specificity. However, these are only two components of a much more encompassing consideration, that is, criterion validity. Although broadly included in the construct of “validity,” it bears particular attention when evaluating instruments for screening purposes. More specifically, what is being referred to here is the previously mentioned diagnostic or classification accuracy or efficiency, which also includes an instrument’s positive, negative, and overall predictive powers. An instrument’s positive predictive power (PPP) indicates the percentage of patients classified as having the disorder or characteristic assessed by the instrument who actually do have that disorder or characteristic. An instrument’s negative predictive power (NPP) indicates the percentage of patients classified as not having the disorder or characteristic assessed by the instrument who actually do not have that disorder or characteristic. Overall predictive power represents the total percentage of those administered the instrument who were accurately classified as having or not having the disorder or characteristic. Diagnostic efficiency statistics are discussed in more detail in Derogatis (2017).

The previously discussed considerations in this and the previous section pertain to the selection of all psychometric instruments regardless of whether the patient is a child, adolescent, or adult. Others have addressed considerations pertaining to the selection of screening measures for pediatric populations specifically. For example, Weitzman and Leventhal (2006) present a question checklist based on the work of Bergman (2004). For the most part, the items in this checklist are addressed in the earlier general considerations section. Drotar, Stancin, and Dworkin (2008) specifically address the selection of pediatric developmental screening instruments as well as provide recommendations for the application of specific instruments. The reader is referred to Drotar et al. for more information, as developmental screening instruments are not addressed in this handbook.

Instrumentation for Treatment Monitoring

Tests that are used for ongoing monitoring of treatment progress over time require considerations that are related to the fact that the instrument will be completed two or more times during the course of an episode of care. Thus, cost becomes a particularly important factor in the selection of a measure to be used for this purpose.

Next, given that the purpose of repeated testing is to detect change in a patient’s status in one or more areas of functioning, it is important that the selected instrument is sensitive to change that
has occurred over time. Here, good test-retest reliability is a key indicator of this sensitivity. Lambert and his colleagues (Burlingame, Lambert, Reisinger, Neff, & Mosier, 1995) indicate that the minimum acceptable reliability should be about .70, while Smith and Archer (2014) recommend that the test-retest value should be at least .80. Ideally, one would want to see reported reliability coefficients that are based on a demographically relevant “normal” or community sample. It is difficult to tell what a stated reliability coefficient based on patient sample really means. Is a low to moderate reliability on a given measure (based on a patient sample) due to true change in the patient as a result of intervention, or is it a reflection of error variance that is built into the instrument? And what does a high, patient-based reliability mean? Does it indicate that the instrument is insensitive to change, or was it based on a sample of patients who truly did not change as a result of intervention? From this author’s perspective, one should feel confident with instruments with good community sample–based reliabilities.

In some cases, using a test with alternate forms might be advisable. For instance, if there is a likelihood that the results of subsequent administrations of a given test may reflect a memory or learning effect, it would be helpful to have a different set of questions that have been empirically demonstrated to yield comparable results to those of the original set. Unfortunately, except for achievement and ability tests, it would be rare to find frequently used treatment monitoring instruments for which there is a validated alternate form.

**Outcomes Instruments**

Generally, outcomes measurement might be considered an extension or variant of treatment monitoring. It is, in effect, the monitoring of the results of treatment at or beyond the end of treatment. For this reason, those additional test-selection considerations for treatment monitoring also apply when tests are being considered for use in outcomes measurement.

Probably the most thorough and clinically relevant guidelines for the selection of instruments for outcomes measurement purposes comes from the National Institute of Mental Health (NIMH)-supported work of Ciarlo, Brown, Edwards, Kiresuk, and Newman (1986). A synopsis of Newman et al.’s (2004) updated summary of this work is presented here.

Newman et al. (2004) describe 11 criteria for the selection of outcomes assessment instruments, each of which can be grouped into one of five types of considerations. The first consideration is that of **relevance to the target group**. The instrument should measure those problems, symptoms, characteristics, and so forth that are common to the population to whom it will be administered. The more heterogeneous the population, the more chance that modifications will be required and that these will alter the standardization and psychometric integrity of the instrument. Related to relevance is the degree to which the results obtained from the instrument’s administration are independent of the type of treatment that is offered to the population being tested.

The second set of general considerations is that of **methods and procedures** (Newman et al., 2004). Several selection criteria are related to this group. The first is that administration of the instrument is simple and easily taught. Generally, this is more of an issue with clinician rating scales than with patient self-report measures. In the case of rating scales, concrete examples or objective referents at each rating level should be provided to the user. This is Newman et al.’s second criterion in this group. The third criterion is that the instrument should allow input not just from the patient but also from other sources (e.g., the clinician, collaterals). This is particularly beneficial in the case of pediatric patients. The benefits of this include the opportunities to obtain a picture of the patient from many perspectives, to validate reported findings and observations, and to promote honesty in responding from all sources (given that all parties will know that others will be providing input also). The fourth methods and procedures criterion is that the instrument provide information relevant to understanding how the treatment may have effected change in the individual.
The last methods and procedures criterion, if adhered to, would significantly limit the number of instruments that could be used for outcomes assessment. It does not appear to be critical for use in integrated pediatric primary care settings at this point in time. Consequently, one may not wish to place too much weight on its consideration unless there are specific reasons for doing so.

Newman et al.’s (2004) third set of considerations has to do with the psychometric strengths of the instrument. Consistent with what was stated earlier, the NIMH panel of experts indicated that outcomes measures should (a) meet the minimum psychometric standards for reliability (including internal consistency, test-retest reliability, and, as appropriate, interrater reliability) and validity (content, construct, and concurrent validity); (b) be difficult to “fake bad” or “fake good;” and (c) be free from response bias and not be reactive or sensitive to factors unrelated to the constructs that are being measured (e.g., physical settings, behavior of the treatment staff).

The fourth group of considerations concerns the cost of the instruments. Newman et al. (2004) point out that how much one should spend on assessment instrumentation and associated costs (e.g., staff time for administering, scoring, processing, and analyzing the data) will depend on how important the data gathered are to assuring a positive return on the functions they support. In the context of the NIMH undertaking, Newman et al. felt that the data obtained through treatment outcomes assessment would support screening/treatment planning, efforts in quality assurance and program evaluation, cost containment/utilization review activities, and revenue generation efforts. However, this should probably be considered the ideal. At this point, the number and nature of the purposes that would be supported by the obtained data will depend on the individual pediatric primary care setting. The more purposes the data can serve, the less costly the instrumentation is likely to be, at least from a value standpoint. In terms of actual costs, Ciarlo et al. (1986) originally estimated that 0.5% of an organization’s total budget would be an affordable amount for materials, staff training, data collection, and processing costs related to outcomes assessment. At the same time, one must be mindful that this recommendation was made more than 30 years ago and may not reflect changes in policies, regulatory and accreditation requirements, rate of inflation, and attitudes related to the use of psychological test instruments since that time.

The final set of considerations in instrument selection has to do with the utility of the instrument. Newman et al. (2004) posit four criteria related to utility. First, the scoring procedures and the manner in which the results are presented should be comprehensible to all with a stake in the treatment of the organization’s patients. This would include not only the patient, their parents, the organization’s administrative staff, and other treatment staff, but also third-party payers and legislative and administrative policy makers. Related to this is the criterion that the results of the instrument are easily interpreted by those with a stake in them. Another utility-related criterion is that the instrument be compatible with a number of clinical practices and theories that are employed in the behavioral health care arena. This should allow for a greater range of problem applicability and greater acceptance by the various stakeholders in the patient’s treatment.

Also important to consider with regard to utility is that “the assessment instruments should support the clinical processes of a service with minimum interference” (Newman et al., 2004, p. 209). There are two issues here. The first has to do with whether the instrument can support the screening, planning, and/or monitoring activities in addition to the outcomes assessment activities. In other words, are multiple purposes served by the instrument’s results? The second issue has to do with the extent to which the organization’s staff is burdened with the collection and processing of assessment data. How much will the assessment process interfere with the daily workflow of the organization’s staff? Equally important is whether the benefits that accrue from the use of the instrument justify the cost of implementing an outcomes assessment program for whatever purpose.

Although the work of Newman and his colleagues provides more extensive outcomes instrumentation selection guidelines than most, others who have addressed the issue have arrived at
recommendations that serve to reinforce and/or complement those found in the NIMH document. Interested readers are referred to the work of Lambert and his colleagues (Lambert & Hawkins, 2004a, 2004b; Ogles et al., 2002), Andrews, Peters, and Teesson (1994), and Burlingame et al. (1995). Overall, one would be wise to heed Ogles et al.’s (2002) three overarching recommendations when evaluating psychological measures for outcomes assessment purposes: know the trade-offs involved in the selection of one measure over another, know the audience, and recognize resource limitations.

Selection of Technology for Psychological Assessment

Maheu, Pulier, McMenamin, and Posen (2012) noted, “Two potent forces—the technological expansion of health care delivery and heightened consumer expectations—present psychologists with the imperative to efficiently and ethically leverage new tools and systems to enhance practice, research, education, training, and policy” (p. 613). Moreover, being able to conduct psychological testing and assessment from outside of the practice setting via telehealth technologies enables the provision of otherwise unavailable services to certain patients and results in convenience and cost and time savings for some patients (Luxton, Pruitt, & Osenbach, 2014).

Looking back to the mid-1980s and early 1990s, the cutting-edge technology for psychological testing included desktop personal computers linked to optical mark reader (OMR) scanning technologies. There also were those “little black boxes” that facilitated the per-use sale and security of test administration, scoring, and interpretive reports for test publishers while making in-office, computer-based testing available and convenient to practitioners. As has always been the case, someone had the foresight to develop applications of current technological advances that we use every day to the practice of psychological testing. Just as at one time the personal computer held the power of facilitating the assessment process, the Internet, FAX, interactive voice response (IVR), and smartphone technologies were later developed to make the assessment process easier, quicker, and more cost-effective. These technologies support the increasing use and popularity of telehealth, “the use of electronic and communications technology to accomplish health care over distance” (Jerome et al., 2000, p. 407). In addition to assessment, telehealth can enable or support the provision of other clinical services, including crisis response, triage, treatment planning, care management, and even psychotherapy.

Perhaps more than ever, the selection of the type of technology by which psychological tests will be administered, scored, and/or interpreted in a pediatric primary care setting requires careful consideration—at least as much as when selecting psychological test measures. Following is a brief overview of technologies that are available to facilitate psychological testing in primary care and other health care settings. This is followed by a discussion of factors that one should consider in the selection of assessment technology for their practice setting.

Common Modalities for Psychological Testing

Following are some of the most current modalities for administering, scoring, and/or interpreting psychological tests.

Paper-and-Pencil Administration Forms

Paper-and-pencil test forms represent the simplest of administration modalities. They have the advantage of not being dependent on the patient having access to a particular device, allowing for the administration of a test just about anywhere and at any time. Commonly, either the indicated responses on the answer sheet can be hand scored or the patient’s responses can be entered into an automated scoring program via key entry or, using specially developed answer forms, via
scanning or faxback technology (see what follows). Moreover, the paper-and-pencil answer form can be used to record answers to test items presented via a validated audiotape version of the test.

**Mail-Out/Mail-Back Administration**

One way of utilizing paper-and-pencil administration is through a mail-out/mail-back process. This administration modality can provide a means of monitoring or assessing the outcomes of treatment for those without access to other available electronic technologies and without needing an in-office visit. It can also serve as a means of screening for a specific psychopathology (e.g., depression) or general psychological distress, or as a means of establishing a baseline prior to an initial office visit.

**Internet Technology**

An Internet-based assessment process is straightforward. The clinician accesses the website on which the desired instrumentation resides. The desired test is selected for administration, and then the patient completes the test online. There may also be an option of having the patient complete a paper-and-pencil version of the instrument and then having administrative staff key-enter the responses into the program. Alternately, the patient may be provided a link to the test administration website, allowing him or her to complete the test outside of the primary care office. The data is scored and entered into the website’s database, and a report is generated and transmitted back to the clinician through the web. Turnaround time for receiving the report can be only a matter of minutes. The test data can later be archived and subsequently be used for any of a number of purposes, such as regularly scheduled reporting of aggregated data to the primary care practice or the psychologist. Data from repeated testing can be used for treatment monitoring and report-card generation, as well as for psychometric test development or other statistical purposes.

The advantages of an Internet-based assessment system are clear. It allows for online administration of tests, which can include branching or other logic for item selection. Instruments available through a website can be easily updated and made immediately available to users. This is in contrast with disk-distributed software, for which updates and fixes are sometimes long in coming. The results of an Internet-based test administration can be made available almost immediately. In addition, data from multiple sites can easily be aggregated and used for the purpose of normative comparisons, test validation, risk adjustment, benchmarking, generation of recovery curves, and any number of other statistically based activities that require large data sets. Moreover, the Internet allows one to access psychological measures that utilize item response theory (IRT) and computerized adaptive testing capabilities (Narayana & Wong, 2015).

There are only a few disadvantages to an Internet-based system. One disadvantage has to do with the general Internet data security issue. With time, this will likely become less of a concern as advances in Internet security software and procedures continue to take place. Moreover, as Tsacoumis (2000) noted, Internet-based testing can be costly, especially when one takes into account fees for Internet access, test administration scoring and reporting services from a reputable publisher or vendor, and equipment maintenance and upgrading. How costly this may become will vary from one provider or organization to another, depending on the particular services and vendors that are used.

**Smartphone Technology**

The commonly found smartphone provides another means of online administration and response entry. After that, entered data are transmitted to the test vendor for scoring, after which the results can be returned to the smartphone for review by the clinician. The real advantage of the smartphone
lies in the ability to conduct an automated administration of a test just about anywhere. Also, it is a device that the vast majority of the U.S. general population is familiar with and is generally well accepted (Clough & Casey, 2015).

**Faxback Technology**

The development of facsimile and faxback technology that has taken place during the past few decades has made available an important application for psychological testing. The process for implementing faxback technology is fairly simple. A paper-and-pencil answer sheet developed for a test available through the faxback system is completed by the patient. In those systems in which several tests are available, the answer sheet for a given test contains numbers or other types of code that tell the scoring and reporting software which test is being submitted. When the answer sheet is completed, it is faxed in—usually through a toll-free number that the scoring service has provided—to the central scoring facility, where the data is both scored and entered into a database. The system may include an editing capability that will allow for the correction of problematic respondent entry errors (e.g., stray marks, multiple responses to single-response items) before final processing of the faxed-in data. A report is generated and faxed or e-mailed back to the clinician, usually within a few minutes. At the scoring end of the process, the whole system remains paperless. Later, the stored data can be used in the same ways as that gathered by an Internet-based system.

Like Internet-based systems, faxback systems allow for immediate access to software updates and fixes. They also can incorporate the same statistical and benchmarking capabilities. Like the PC-based testing products that are offered through most test publishers, their paper-and-pencil administration format allows for more flexibility as to where and when a patient can be tested. In addition to the types of security issues that come with Internet-based testing, the biggest disadvantage of or problem with faxback testing centers around test identification and linkage of data obtained from an individual patient. Separate answer sheets are required for each instrument that can be scored through the faxback system. The system must also be able to link data from multiple tests or multiple administrations of the same test to a single patient.

An example of how faxback technology can be implemented for screening purposes in primary care settings can be found in a project described by Goldstein, Bershadsky, and Maruish (2000).

**IVR Technology**

Another application of technology to the administration, scoring, and reporting of results of psychological tests can be found in the use of interactive voice response, or IVR, systems. Everyone is familiar with IVR technology. When one calls to order products, address billing problems, find out the balance in their checking account, or conduct other phone-enabled activities, one is often asked to provide information to an automated system, either orally or by key entry using the phone’s keypad, in order to facilitate the meeting of the caller’s requests. This is IVR, and its applicability to test administration, data processing, and data storage is simple.

In essence, IVR test administration is similar to the automated online, desktop administration of psychological measures. However, as Kobak, Greist, Jefferson, and Katzelnick (1996) pointed out, the advantages of IVR over PC-based testing include the patients’ comfort with the technology (i.e., touchtone telephone), easy availability and access to that technology, and the ability of the patient to complete the administration outside of the practitioner’s office at any time of the day or week.

IVR technology’s utility as a test delivery system has been reported in a number of published studies. Kobak and his colleagues conducted several studies that employed the use of IVR technology for administration of various types of psychological instruments. Kobak et al. (1997) found
support for the overall validity and use of both IVR and desktop-computer administered versions of the PRIME-MD as instruments for gathering information from behavioral health patients. Here, the results from the clinician-administered Structured Clinical Interview for DSM-IV for Axis I Disorders (SCID-I)—Clinician Version (First, Spitzer, Gibbon, & Williams, 1997) were used as the diagnostic gold-standard criteria.

In their review of 10 studies that included the administration of both clinician- and IVR-administered versions of the Hamilton Depression Rating Scale (HDRS; Kobak, Reynolds, Rosenfeld, & Greist, 1990), Kobak, Mundt, Greist, Katzelnick, and Jefferson (2000) found the IVR system to have generally yielded comparable or better psychometric properties than the clinician version. The combined data across all 10 studies yielded an overall correlation of .81 ($p < .001$) between the scores resulting from the administration of the HDRS in both formats.

Moreover, in their HDRS and other studies, Kobak et al. (2000) reported that up to 90% of patients being assessed by IVR systems reported moderate to very strong acceptance as it related to clarity and ease of use. These findings, in turn, are consistent with Kobak et al.’s (1997) findings in the PRIME-MD study. Findings from these and other studies led Kobak et al. (2000) to observe that “several decades of research have shown that people often report more problems of greater severity to computer applications than to clinicians, especially when such problems concern personally sensitive topics” (p. 152). Millard and Carver (1999) arrived at similar findings in their comparison of results obtained from the administration of the SF-12 Health Survey via the IVR and live telephone interview formats.

IVR technology is attractive from many standpoints. It requires no extra equipment beyond a touchtone telephone or cell phone for administration. It is available for use 24 hours a day, 7 days a week. One does not have to be concerned about the patient’s reading ability, although oral comprehension levels need to be taken into account when determining which instruments are appropriate for administration via the IVR (or other audio) administration format. As with faxback- and Internet-based assessment, the system is such that branching logic or a more sophisticated item selection system (e.g., IRT) can be used in the administration of the instrument. Updates and fixes are easily implemented systemwide. Also, the ability to store data allows for comparison of results from previous testings, aggregation of data for statistical analyses, and all the other data analytic capabilities available through faxback- and Internet-based assessment.

In addition, Kobak et al. (2000) identified several methodological benefits of IVR. Among them are the provision of a completely standardized, structured assessment; thorough and complete error checking of data at the time of collection and database entry; immediate real-time access to analyzable data; patient-determined pace of assessment; and increased honesty on sensitive questions.

As for the down side of IVR assessment, probably the biggest issue is that in many instances, the patient must be the one to initiate the testing. This is a particular concern when the patient is a child or adolescent. Thus, control of the testing is turned over to a party that may or may not be interested in or otherwise amenable to psychological testing. With less cooperative patients, this may require costly follow-up efforts to encourage full participation in the process.

**General Considerations for Technology Selection**

As noted, each of the previously described assessment technologies has its own set of advantages and disadvantages. How does a primary care setting determine which technology is best for its particular needs? For the individual pediatric primary care psychologist, there may not be a choice. The primary care setting in which the practitioner provides services may have an assessment system in place—instrumentation, delivery system, and associated reports—and the practitioner may be required to use it. Alternately, the practitioner may be free to employ any test delivery system that best meets their needs. Regardless of whether the decision maker is the
psychologist or someone else, several points should be considered when selecting from among
the available technologies.

**Psychometric Integrity of Automated Test Administration**

Luxton et al. (2014) warn that those conducting psychological assessments via telehealth tech-
nologies must be aware of the psychometric properties of tests that are administered using those
technologies, particularly when test administration takes place outside of the clinician’s office. For
example, lack of an in-person presence may have an effect on the patient’s clinical presentation
and honesty in responding—again, a very important consideration with pediatric patients. The
technology being employed may be inadequate for the test administration task. Poor network con-
nnections, bandwidth limitations, equipment-related distractions, and patient fatigue or discomfort
(e.g., eyestrain from viewing the monitor) are important considerations when evaluating the test’s
online psychometric properties. Moreover, patient acceptance of the employed technology as well
as cultural and demographic factors must be taken into account when evaluating technologies for
psychological test administration.

Noting that the literature reports evidence of both equivalence and differences between paper-
and-pencil and Internet administered standardized measures, Luxton et al. (2014) concluded:

> there are gaps in the literature that practitioners should consider when selecting particular
> assessment instruments and [technology] mediums. In particular, the vast majority of available
> measures and assessment tools are based on norms that were established by employing tradi-
> tional in-person procedures. The reevaluation of these tools with diverse populations, clinical
> presentations, and telehealth mediums is necessary to assure the validity of assessments con-
> ducted via telehealth technologies. It is critical for practitioners to be cognizant of assessment
> measure limitations and to appropriately disclose and document them in their practice.

*(p. 32)*

**Cost**

Just as it is for test instrumentation, the cost of a system that will administer, score, store, and/or
report the results of psychological measures is a major consideration. In addition to any up-front
equipment purchase or development costs, one must also consider the associated costs. Depending
on the delivery system, additional expenses may result from software licensing fees, ongoing
technical support, maintenance contracts, system upgrades, adding new measures to the system,
and ongoing phone line and/or Internet provider fees, just to mention a few. These costs will
vary from one primary care setting to another, depending on the resources that may already be
available to the purchaser.

**Availability/Accessibility**

The technologies discussed earlier may not be available to or accessible by all involved par-
ties. Good examples here are smartphones and tablets. Although it seems that everyone has a
smartphone, this certainly is not the case. According to survey conducted in 2015, 73% of U.S.
adults own a desktop or laptop computer and 68% have a smartphone, but only 45% own a tablet
computer (Pew Research Center, 2015). A primary care setting wishing to initiate an off-site
smartphone- or tablet-based assessment system should therefore determine what percentage of
their patient population has access to a smartphone or tablet before completing the decision-
making process. Faxback and IVR might be less problematic for providers and/or patients to
access, but this should not be assumed without evidence to this effect.
Compatibility with Existing EHR Initiatives

Electronic health record (EHR) systems’ contribution as a tool for health care improvement has been recognized and continues to be adopted by health care organizations (Kobus, Harman, Do, & Garvin, 2013; Tsai & Bond, 2008). The existence of over 2,400 EHR vendors in 2014 (Mandros, 2015) attests to the acceptance of and demand for these systems. The utility of their use with psychological test data and other behavioral health treatment data has been investigated in several studies (for example, see Gill, Chen, Grimes, & Klinkman, 2012; Gill & Dansky, 2003; Klein, Hunt, & LeBlanc, 2006; Rollman et al., 2002). Such studies show that acceptance of these data into EHRs can facilitate such things as a clinician’s quick access to screening results at the time of the patient visit, tracking patient results over time, and generation of periodic reminders for patient follow-up.

With regard to the present discussion, one needs to evaluate the degree to which the psychological testing technology being considered for the primary care practice is compatible with the current or planned EHR and other information technology (IT) systems that support the practice. It is unlikely that the choice of an EHR or other IT for a primary care practice will be heavily based on the psychologist’s preferred psychological testing system. Thus, the key question is which mode of administration, scoring, and data entry is most compatible with the practice’s existing or anticipated EHR or IT system. For example, although an IVR system may be appealing to a primary care setting for many reasons, the practice’s plans to develop a multifunction website would suggest that an Internet-based test delivery system would be more practical and cost-effective, all other considerations being equal. For a small primary care group practice, one may find faxback, IVR, or Internet solutions to be equally viable.

Ease of Implementation

Regardless of who the decision maker is, one must consider how easy it will be to implement the assessment delivery and reporting technology in daily clinical practice. How receptive will the clinical and support staff be to the selected technology? How easy will it be to train them in its use? What demands will its use make on them? Where will its use be inserted into the daily workflow? How receptive will patients be to this technology? How difficult will it be for patients to understand how to complete an assessment instrument via this technology? These all are important questions and should have a strong bearing on any decision related to the selection of technology for the purpose of test administration, scoring, and reporting. Especially in the case of outcomes assessment, the success of an outcomes program can hinge on the degree to which a workable plan of implementation—including technology-related aspects—can be developed.

Patient Affordability

Situations in which testing is not fully covered by an insurer make the costs incurred by the patient worthy of careful consideration. This is particularly true if there is a per-use fee for the instrument(s) being administered and a separate charge for the use of the automated system. The full or co-pay charge incurred by the patient’s parents will be made more acceptable if they can see value in both the need for testing and the use of the system. If the benefits of the test delivery are not obvious to the parent or guardian, this may impact the relationship with the provider.

Access to Measures Selected for Use

No matter how sophisticated the technology is, it will be of limited utility if the instrumentation that the psychologist or primary care practice wishes to use is not available or cannot be developed
Logistics of Readministration of Measures

This becomes an issue when psychological testing is part of a system of treatment monitoring or outcomes assessment. Obtaining a baseline measure of a patient’s status is not that difficult when he or she is in office. At intake, patients generally are more amenable to completing psychological test instruments. Either later during treatment or after the episode of care has been completed, the interest or willingness to complete the same psychological measure one or more times again is usually on the wane.

A provider’s monitoring of patient progress while the treatment is ongoing is facilitated by the fact the patient can be retested while he or she is in office. After treatment has been completed, the ability to conduct a follow-up remeasurement for outcomes management purposes becomes more difficult, since the primary care practice no longer has ongoing regular or scheduled contact with the patient. Thus, there must be a feasible means of conducting a “long-distance” follow-up assessment of the patient. This would include being able to cost effectively contact the patient for retesting, provide him or her with a means of easily completing the measure(s) in question, and having the data returned for analysis. For some, this problem may be solved through the use of an IVR system. As was described earlier, the primary care practice would send the patient or their parent a letter at a specific time following the termination of treatment, asking the patient to retake the baseline measure. Included in the letter would be a toll-free, phone-in number and personal identification number to access the system. One might increase the response rate by also enclosing a written copy of the instrument and giving the patient the option of using the IVR system or completing the paper form and mailing it back. A similar method might be used if the primary care setting’s Internet assessment system is accessible to patients. This process may be difficult with pediatric populations owing to the patients’ age.

Flexibility

Flexibility refers to the degree to which any number of test administration and reporting options are available through the technology. For example, faxback and IVR technologies only allow for data input via one means, that is, through either completion and faxing of a paper-and-pencil form or dial-in and telephone keypad entry, respectively. Internet-based testing (along with in-office PC-based assessment) allows for online administration and data entry, as well as key-entry or, in some cases, optical scanning of data gathered from a scannable pencil-and-paper test form. Flexibility will not be an important consideration in many cases. However, in those cases in which maximization of data collection or meeting a minimum quota of tested individuals is critical (e.g., surveying service satisfaction of a specified minimal number of members from a given health plan), having a way to gather data by more than one means can help increase the chances of success in the endeavor.

Clinical Utility

In Gray’s (1999) discussion of factors to consider in selecting a clinical outcomes system, he addresses clinical utility. In this context, Gray refers to clinical utility as “the ability to provide
the clinician with real-time feedback” (p. 9). Thus, one should ascertain how quickly assessment information is accessible once the patient completes the measure. This can be an extremely important consideration in evaluating systems that will be used to support screening, treatment monitoring, and other clinical activities involving real-time, in-office clinical decision making. Information that is not immediately available to the clinician may prove to be of little or no value.

Although important, Gray’s (1999) view of clinical utility is quite limited. One may extrapolate from earlier discussions in this section and those of Fischer and Corcoran’s (1994) and Newman et al.’s (2004) recommendations pertaining to the evaluation of the clinical utility of psychological instruments in order to consider other important factors when evaluating technologies and modalities to use to conduct psychological assessments. Included here are factors such as patient acceptance, ease of use by all parties involved (e.g., patients, parents, clinicians, support staff), compatibility with the practice’s EHR, and the available technology’s ability to support multiple assessment activities (e.g., screening, treatment monitoring, outcomes assessment). Finally, one must also consider the technology/modality’s cost-effectiveness and its overall advantages over other options when judging its clinical utility.

Summary
Choosing the right measures and technology for psychological screening or assessment conducted in a primary care setting can be a difficult task. This chapter has tried to convey various criteria, features, issues, and concerns that the author and other recognized experts feel are important to consider in selecting psychological measures and the technology to support their use in pediatric primary care settings. In reality, there is no one set of criteria that either instrumentation or technology must or should meet. Instead, decision makers must decide what to measure, why they want to measure it, and how they plan to use the resulting information within their specific pediatric primary care practice. Based on the answers to these questions, determining the importance and necessity (if any) of each of the considerations presented in this chapter as it pertains to the pediatric primary care setting’s goals or the psychologist’s needs will be a relatively simple task.

Author’s Note

References


Psychological Measures and Reporting Technologies


