Evidence-Based Assessment of Child and Adolescent Disorders: Issues and Challenges

Eric J. Mash
Department of Psychology, University of Calgary

John Hunsley
School of Psychology, University of Ottawa

The main purpose of this article and this special section is to encourage greater attention to evidence-based assessment (EBA) in the development of a scientifically supported clinical child and adolescent psychology. This increased attention is especially important in light of (a) the omission of assessment considerations in recent efforts to promote evidence-based treatments for children and (b) ongoing changes in the nature of clinical child assessment. We discuss several key considerations in the development of guidelines for EBA, including the purposes of assessment, the role of disorder or problem specificity, the scope of assessment, assessment process parameters, possible "cross-cutting" assessment issues, psychometric considerations, and issues related to the clinician's integration of assessment data. We conclude the article with suggestions for how current, summary information on EBA can be developed, maintained, and disseminated.

Measurement is one of our most ordinary actions. We speak its language whenever we exchange precise information. ... This very ubiquity, however, makes measurement invisible. ... So it is not surprising that we take measurement for granted and consider it banal. Yet the use a society makes of its measures expresses its sense of fair dealing. ... Our methods of measurement define who we are and what we value. (Alder, 2002, pp. 1–2)

It would be difficult to imagine providing any form of psychological services to children and families without some type of informal or formal assessment, at a minimum to assess whether services should be provided or a referral made. Although details may vary across theoretical perspectives, virtually all definitions of psychological assessment refer to (a) assessment processes (i.e., generating and testing hypotheses, case formulation, decision making, integration of information) and (b) assessment methods and measures (i.e., the use of tests, questionnaires, observations, and other measures to gather data), which are used to answer a variety of questions about a child and to develop and evaluate appropriate interventions (Fernandez-Ballesteros et al., 2001; Mash & Terdal, 1997a; Ollendick & Hersen, 1993).

Many clinical psychologists view assessment as a unique and defining feature of their expertise in the context of multidisciplinary child healthcare settings and services (Krishnamurthy et al., 2004). In fact, assessment is such an integral component of clinical child psychology practice that its value in support of this endeavor is typically assumed. However, solid evidence to support the usefulness of assessment for improving treatment outcomes for children who are assessed is lacking, and many assessment methods com-
EVIDENCE-BASED ASSESSMENT

monly used with children are not supported by the evidence (see Fletcher, Francis, Morris, & Lyon, 2005; Neisworth & Bagnato, 2004). As Peterson (2004) recently noted, "For many of the most important inferences professional psychologists have to make, practitioners appear to be forever dependent on incorrigibly fallible interviews and unavoidably selective, reactive observations as primary sources of data" (p. 202). This state of affairs has existed for a long time, leading some clinicians to question whether quantification of important clinical child practice constructs, processes, and outcomes is even possible (Garland, Kruse, & Aarons, 2003).

Of perhaps even greater concern than the lack of evidence for the contribution of assessment to treatment outcome is that such evidence is rarely, if ever, sought or obtained. This pervasive neglect reflects several interrelated factors, including (a) a steadfast belief in the intrinsic worth of many established and habitually used assessment methods (e.g., intelligence tests and projective personality tests); (b) difficulties in operationalizing the multifaceted and iterative parameters of assessment, which, as we noted, encompass both processes and methods; (c) a lack of clarity regarding the kinds of evidence and criteria needed to evaluate the effectiveness of assessment methods and processes; (d) factors that conspire to make assessment activities somewhat less "glamorous" than treatment or intervention activities, particularly their (rightful) role as an ancillary activity or intermediate step in the service of effective treatment; and (e) diminishing availability of health care resources for assessment activities, resulting in a negative cycle in which a lack of documented usefulness leads to a reduction in assessment resources, resulting in a further reduction in seeking evidence that might support the potential benefits of assessment.

As a result of these and other factors, it is rare that assessment processes such as the way in which clinicians integrate and use assessment information in case formulation, treatment planning, or treatment monitoring are ever evaluated (Fernandez-Ballesteros et al., 2001). Accordingly, although cognizant of the importance of evaluating assessment processes, the absence of data in this area leads to an emphasis on assessment methods in this article and, indeed, in this special section. Although assessment methods are evaluated frequently, these evaluations focus almost exclusively on traditional psychometric criteria (e.g., reliability, validity) as described in test manuals, rather than on the methods' demonstrated applied value for the purposes and clinical populations they are intended to serve (cf. Hunsley, 2003; Mash & Hunsley, 2004). In the current climate of evidence-based child mental health practice (Ollendick & King, 2004), systematic efforts to consider what constitutes evidence-based assessment (EBA), both in general and in relation to specific child disorders and problems, seem long overdue (Hirsh-Pasek, Kocharoff, Newcomer, & de Villiers, 2005). This article and the articles in this special section that follow explore several key considerations in the development of guidelines for EBAs needed to assist clinicians, administrators, and policymakers in providing mental health services to children.

The present state of neglect with respect to EBA is not unlike what existed 10 years ago for evidence-based treatments (EBTs) for children and adolescents (Kazdin, 2000). However, the rapid growth in the development and promotion of EBTs in clinical child and pediatric psychology over the past decade (e.g., Barrett & Ollendick, 2004; Chorpita et al., 2002; Herschell, McNeil, & McNeil, 2004; Hibbs & Jensen, 2005; Journal of Pediatric Psychology, 1999–2001; Kazdin & Weisz, 2003; Longan, Elbert, & Bennett-Johnson, 1998; Mash & Barkley, in press; Weisz, 2004) has not been matched by similar efforts to develop and use EBAs (Hunsley & Mash, in press; Ollendick, 2003). Despite repeated calls for greater attention to EBA (e.g., Achenbach, 1985; Achenbach & McConaughy, 1997; Achenbach & Rescorla, 2001; Mash & Torday, 1997a; Ollendick & Hersen, 1984), there have been few concerted efforts to develop criteria for the selection, use, and evaluation of processes and methods for the assessment of child and adolescent disorders (for an exception, see American Psychological Association, Division 12, Task Force for Upgrading the Science and Technology of Assessment and Diagnosis; Frick, 2000). It seems self-evident that assessments should be "useful" in designing and evaluating effective and efficient services for children and families. In fact, "an understanding of the relationship between assessment and intervention," "intervention planning," and "the ability to assess outcomes of treatment/intervention" have all been identified as core competencies and qualifications associated with psychological assessment (Krishnamurthy et al., 2004, p. 732; see also American Psychological Association, 2000). Nevertheless, although their importance is widely acknowledged, the nature and strength of the links between assessment and intervention remain tenuous at best, and the role of assessment in EBT is virtually unexplored (Weisz, Chu, & Polo, 2004).

Given this lacuna between EBT and EBA, it is rather ironic that EBTs can be identified as evidence-based only on the basis of solid assessment data (Silverman & Saavedra, 2004), which are often lacking or are based on traditional diagnostic approaches or conventional tests that may have little direct relevance for treatment (Bickman, 2002; Neisworth & Bagnato, 2004). For the most part, treatment-focused task force statements and guidelines have been silent on the use of assessment in developing and evaluating EBTs, notwithstanding general recommendations to use a "reliable and valid" core assessment battery— recommen-
dations that are presented with little guidance as to what makes an assessment method or process reliable or valid in a particular context or for a particular purpose. Assessment methods are used for a multitude of purposes; thus, given the conditional nature of test reliability and validity, fine discriminations must be made to determine when reliability and validity coefficients found for one assessment purpose or group are generalizable to other purposes or groups (Haynes, Richard, & Kubany, 1995). Consequently, blanket recommendations to use reliable and valid measures when evaluating treatments are tantamount to writing a recipe for baking hippopotamus cookies that begins with the instruction “use one hippopotamus,” without directions for securing the main ingredient.

The impetus for this special section on EBA with children and adolescents derives from the need for an assessment equivalent to the EBT task force statements on what constitutes a sufficient evidence base for psychological interventions. Compared to EBT initiatives, the task of developing and documenting EBAs is a far greater challenge due to (a) the sheer number of assessment methods and processes for particular problems and outcomes relative to the number of available treatments and (b) the many purposes of assessment as compared with treatment. This challenge is compounded in assessments of children, where developmental changes in the domains being assessed (Lahey et al., 2004) and the embeddedness of children in the family and peer group require that a much larger number and variety of methods be developed and used than is the case for adults. For example, compared to adult assessments, it is much more common in youth assessments to seek information from multiple informants and to incorporate into the assessment data from naturalistic observations (e.g., school observations, home observations). Adding to this challenge is that child assessment by its very nature involves the use of ongoing decision-making processes, often requiring an integration of information obtained at different ages, from repeated assessments (before, during, and following treatment), using multiple methods (e.g., interviews, ratings, direct observations), informants (e.g., child, parent, teacher), and settings (e.g., home, classroom), and not just the utilization of a collection of psychometric methods (Evans & Meyer, 1985; Mash & Terdal, 1997a). Any attempt at delineating EBA must be sensitive, therefore, to the multifaceted nature of assessments with children and families. Our main goal in this introductory article is to help redress the relative inattention to assessment in the development of an evidence-based clinical child psychology practice by discussing a number of key conceptual and pragmatic issues that must be considered in the development of standards for EBA of child and adolescent disorders and in the promotion of EBA more generally. We use the term EBA to describe assessment methods and processes that are based on empirical evidence in terms of both their reliability and validity as well as their clinical usefulness for prescribed populations and purposes. The term evidence-based has become commonly associated with professional health care practices that are based on an amalgamation of scientific data, clinical expertise, and patient preferences (American Psychological Association, 2005; Institute of Medicine, 2001; Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000). In contrast, similar terms such as empirically validated or empirically supported have been contentious (see Lampropoulos, 2000, for a summary of some of these concerns) and are identified primarily with treatment task force reports.

In the sections to follow, we consider the need for EBA guidelines, the changing nature of psychological assessment, the multiple purposes of assessment, the importance of disorder or problem specificity, the scope of assessment, assessment process parameters, possible “cross-cutting” assessment issues, psychometric considerations, and issues related to the clinician’s integration of assessment data. Rather than proposing a set of absolute criteria that need to be met for a measure to “count” as evidence based, we provide an initial set of dimensions that could be used when examining assessment methods and some suggestions on the criteria that might be considered in profiling the level of evidence available for these dimensions.

Are EBA Guidelines Needed?

It could be argued that the development of explicit guidelines for EBA are unnecessary given that clinical child psychologists, whether trained in the scientist-practitioner or local clinical scientist model, by virtue of their training should already possess sufficient knowledge and skills to ensure the use of EBA options (Elbert, 1985; Hunsley, Crabb, & Mash, 2004). However, the research literature on child and family assessment and assessment-relevant domains is so voluminous as to render the task of remaining current with the relevant evidence almost impossible for most practicing psychologists. Similarly, the plethora of new or revised methods to assess the same or similar constructs that are available from commercial publishers and presented in scientific journals makes the task of selecting optimal methods based on psychometric and other criteria—and reevaluating the use of these methods on a regular basis—time consuming and costly for most clinicians (Haynes & Lench, 2003), assuming that they even have access to the necessary scientific resources. Thus, the development and availability of EBA guidelines should be a significant boon to clinicians. Such guidelines are intended to aid clinicians not to replace
their judgments or decisions; they are intended to orient clinicians to relevant factors in service provision while recognizing that the ethical, professional, and legal responsibility for service delivery choices and implementation lies with the individual clinician and service delivery system.

A key question is whether currently available assessment methods have clinical utility (Hayes, Nelson, & Jarrett, 1987; Nelson-Gray, 2003). That is, do they provide psychologists with the kinds of information that can be used in ways that will make a meaningful difference in relation to diagnostic accuracy, case formulation considerations, and treatment outcomes? The clinical utility of assessment is also concerned with assessment-related costs, improvements in usual clinical decision making due to the assessment, changes in rates of false positives and false negatives associated with the assessment (based on sensitivity and specificity indexes), and the economic and psychological costs associated with these errors (Hunsley, 2003). Similar to the differences that have been described in research-based versus clinic-based therapy (Connor-Smith & Weisz, 2003; Weisz, Donenberg, Han, & Weiss, 1995), gaps exist between the applicability and use of assessments developed and used in research versus real-world settings. In many instances, measures developed for research may be too long or costly for clinical use. Even when EBA practices are disseminated using structured implementation strategies, they may not change clinicians’ behaviors (Croudace et al., 2003). One survey, for example, found that even when outcome assessments were mandated in clinical practice and training was provided in their use, most clinicians reported never using the obtained scores in their practice because of the time burden or concerns about the scope and validity of the measures, instead relying primarily on clinical intuition in their assessments (Garland et al., 2003). Other surveys have found that about half of psychologists report “sometimes” using formal assessments to evaluate treatment outcomes, and nearly 20% report “never” doing so (Cashel, 2002).

Of course, such surveys are subject to a variety of selection and reporting biases, and the ways in which assessment methods are selected and used by psychologists in applied settings, particularly in relation to treatment planning and outcome evaluation, remain largely unknown. Nevertheless, there is extensive evidence that many clinical child psychologists (a) are trained in and continue to use some measures for which there is no supporting evidence (Belter & Piotrowski, 2001; Cashel, 2002; Hunsley et al., 2004), (b) use these measures in a manner that is uninformed by the documented limitations of the measures (Garb, 1998; Hunsley, Lee, & Wood, 2003), and (c) may underutilize cost-effective evidence-based measures (Palmiter, 2004).

The Changing Nature of Psychological Assessment

In the past two decades, the essence of psychological assessment has changed dramatically, concomitant with a significant decline in the amount of time psychologists devote to assessment activities (Piotrowski, 1999). Influenced by multiple factors, including cost and time considerations (Cashel, 2002; Yates & Taub, 2003), methodological developments in assessment research (e.g., Frick, 2000; Frick & Cornell, 2003; Wood, Garb, Lilienfeld, & Nezworski, 2002), and the dissemination and use of behavioral assessment principles and practices in mainstream assessments (Haynes & Heiby, 2004; Mash & Hunsley, 2004), psychological assessment of children is much more than the use of an intelligence test, an achievement test, one or two projective tests, and one or two measures of psychopathology.

There has been a significant move in clinical child assessment research and practice away from a generic test battery to the use of disorder-specific, brief, face valid, symptom or problem-focused measures, and batteries that are inexpensive and can be integrated into treatment services (Mash & Terdal, 1997a). This newer generation of measures and batteries is of particular interest to health care system administrators who require practical assessment strategies that (a) are appropriate for evaluating services using standardized outcome data, (b) can be used for computer-supported monitoring and continuous quality improvement, (c) can be readily disseminated to community and service organizations, and (d) can be used to develop profiling tools and benchmarking strategies (Eisen & Dickey, 1996; Glisson, 2002; Percevic, Lambert, & Kordy, 2004; Stricker, Troy, & Shueman, 2000).

At present, many general resources and guidelines are available to assist clinicians in identifying assessment practice parameters and in selecting and using specific methods for assessing a full range of child problems or disorders and family functioning (e.g., American Academy of Child and Adolescent Psychiatry, 1997a, 1997b; American Academy of Pediatrics, 2000; American Psychiatric Association, 2000b; Antony, Orsillo, & Roemer, 2001; Division for Early Childhood [Neisworth & Bagnato, 2000]; Fernandez-Ballesteros et al., 2001; Kamphaus & Frick, 2002; Kelley, Noell, & Reitman, 2003; Mash & Terdal, 1997b; Nezu, McClure, Ronan, & Meadows, 2000; Sattler, 2001, 2002; Touliatos, Perlmutter, Straus, & Holden, 2001). No fewer than 10 professional organizations and agencies committed to the welfare of children (i.e., American Academy of Child and Adolescent Psychiatry; American Academy of Pediatrics; National Association of School Psychologists; American Speech, Hearing, and Language Association; American Occupational Therapy Association; Council for Exceptional
Children; National Association for the Education of Young Children; National Institute for Clinical Excellence; Head Start Performance Standards; Society for Pediatric Psychology; and Society for Behavioral Pediatrics) have presented best-practice parameters for a wide variety of childhood problems and disorders in their journals, task force reports, and Web sites (e.g., www.nice.org.uk/page.aspx?o=245600; www.aacap.org/clinical/parameters/;www.effectivechildtherapy.com/), many of which include guidelines for assessment. A number of barriers, however, make it difficult for clinicians to translate these general assessment guidelines into practice, not the least of which are the large number of guidelines that have been proposed; a perceived lack of relevance between guidelines and practice; difficulties in applying guidelines to individual cases; a lack of staff time, knowledge, and resources; a lack of reimbursement for assessment activities; and a fragmented referral system (Leslie, Weckerly, Plemmons, Landsverk, & Eastman, 2004). Even when practice guidelines are available and used, ongoing examination of their validity is needed to determine whether some guidelines need to be withdrawn, retained in a modified form, or retained as originally presented. For example, one review of the validity of 17 previously established medical practice guidelines found that more than 80% needed to be withdrawn or modified (Ortiz, Eccles, Grimshaw, & Woolf, 2002).

The 1999 Standards for Educational and Psychological Testing (also see Code of Fair Testing Practices in Education, 2004) provides a useful framework for ensuring that relevant scientific issues are addressed by test developers and test users (Eignor, 2001). The Standards are intentionally nonprescriptive with respect to the level and quality of scientific evidence needed to ensure that a test is adequate for specific purposes. For example, Standard 2.1 indicates that relevant reliability values and standard errors of measurement should be reported for each subscore, combination of subscores, or total score derived from a test but is silent on such elements as the size and composition of the reliability sample and the optimal level of reliability for clinical use in specific contexts. Accordingly, we believe that the clinical child psychology assessment literature has developed to the point where it is desirable, feasible, and necessary to move beyond the generic framework offered by the Standards. Specifically, the assessment field needs to move toward developing explicit, evidence-based criteria to ensure the reliability, validity, and utility of clinical assessment activities for specific child and adolescent disorders, purposes, and contexts.

In summary, despite the changing nature of psychological assessment and the availability of many excellent resources and guidelines for assessing children and families, there remains a pressing need for evidence-based guidance, drawing on other scientific literatures (such as those in developmental psychology, developmental psychopathology, prevention research, and treatment research), on the assessment methods and processes that are likely to have clinical value in service provision. In the context of a service consisting of assessment only, this would mean focusing on methods and measures that (a) tap key aspects of the presenting problems and associated features as derived from relevant child and adolescent psychopathology research (Mash & Barkley, 2003); (b) have (if at all possible) incremental validity (i.e., a measure adds to the prediction of a criterion beyond what can be predicted with other data (Hunsley, 2003; Hunsley & Meyer, 2003; Johnston & Murray, 2003); and (c) are directly relevant to rendering a diagnosis, generating a case formulation, and generating broad recommendations for treatment. In the context of treatment-related assessments, this would mean using methods that have demonstrated value in the treatment literature for treatment planning, treatment monitoring, and treatment evaluation in relation to specific child and adolescent disorders (Mash & Terdal, 1997b).

**Purposes of Assessment**

An obvious characteristic of psychological assessment with children that must be taken into account in any effort to define EBA is that assessments can be conducted for many different purposes. Recent standards for test users acknowledge the importance of considering both assessment purposes and contexts (American Psychological Association, 2000). Overall, it is possible to identify a small number of interrelated purposes that form the basis for most child assessments. These are (a) diagnosis and case formulation (i.e., determining the nature or causes of the presenting problems, which may or may not involve the use of a formal diagnostic or categorization system), (b) screening (i.e., identifying children who have or who are at risk for a particular problem and who might be helped by further tests or treatment), (c) prognosis and other predictions (i.e., generating predictions about the course of the problems if left untreated, recommendations for possible courses of action to be considered, and their likely impact on the course of the problems), (d) treatment design and planning (i.e., selecting or developing and implementing interventions designed to address children’s problems by focusing on elements identified in a diagnostic evaluation), (e) treatment monitoring (i.e., tracking changes in symptoms, functioning, psychological characteristics, intermediate treatment goals, and variables determined to cause or maintain the problems), and (f) treatment evaluation (i.e., determining the effectiveness, social validity, consumer satisfaction, and cost effectiveness of the intervention).
The conditional nature of test validation requires that the assessment purpose be considered in determining whether a method, and how it is used, can be considered evidence based for that purpose. Although some assessment methods are used for multiple purposes, validity evidence for each major form of assessment purpose must be considered separately. For example, structured interviews such as the Diagnostic Interview Schedule for Children (Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000) and Child and Adolescent Psychopathology Scale (Lahey et al., 2004) may have validity for the purposes of determining the nature of a child’s problem, screening children who are at high risk, or determining that treatment is needed but may have limited validity for the purposes of treatment design and planning.

Accordingly, in considering the establishment of criteria for EBA, it is crucial that they are sensitive to the purposes of the assessment, for decisions regarding the appropriateness or usefulness of particular assessment strategies and measures can only be made in relation to the requirements of the specific assessment purpose or situation. Thus, discussions concerning the relative merits of information obtained via self-report versus direct-observation methods, for example, have little meaning outside of the assessment purpose and context (McFall, 1986). The use of a multivariate observational coding system may be appropriate for assessments designed to identify potential child behaviors and possible maintaining factors to be addressed in treatment but may be insensitive or unnecessary for evaluating the outcome of a highly focused intervention for a specific phobia, where a self-report measure or behavioral avoidance test would likely be more useful and cost efficient (Barrios & Hartmann, 1997).

Considerations of the psychometric characteristics of measures also must be embedded in the purposes of the assessment. Awareness of the importance of specificity, sensitivity, positive predictive power, negative predictive power, and related validity statistics is growing (Hsu, 2002). This family of validity statistics is particularly relevant for diagnostic and prognostic assessment purposes and contains essential information for any measure that is intended to be used for screening purposes. Receiver-operating characteristic (ROC) curves are decision-making tools for determining sensitivity and specificity at various cutoff points for a measure that are independent of base rates of the characteristic under consideration. It may be appropriate, therefore, to require that there are ROC curves available to clinicians for measures intended to be used for screening and diagnostic purposes (McFall & Treat, 1999). ROC curves have recently been described for general inventories of child behavior (Danielson, Youngstrom, Findling, & Calabrese, 2003) and for measures of specific childhood disorders such as depression (Timbremont, Braet, & Dressen, 2004), developmental coordination disorder (Hay, Hawes, Faught, & Hay, 2004), eating disorders (Mond et al., 2004), externalizing disorders (Hudziak, Copeland, Stanger, & Wadsworth, 2004), and anxiety disorders (Dierker et al., 2001).

Some of these validity statistics may have little relevance, however, for many methods intended to be used for treatment monitoring or evaluation purposes. For these purposes, sensitivity to change is a much more salient psychometric feature (Vermeersch, Lambert, & Burlingame, 2000). This involves a determination of whether scores on a measure change in the predicted direction following treatment, whether the change on the measure is significantly greater in treated versus untreated children, and whether the change is not simply due to measurement error. Few measures are designed with treatment sensitivity in mind, and there are still relatively few studies examining the extent to which various child assessment methods are able to reliably detect clinically significant changes in relation to specific forms of treatment.

For some adult problems, there is accumulating evidence that certain assessment methods can be sensitive to treatment effects. The use of role-playing methods in evaluating social functioning is a good example (Norton & Hope, 2001). However, in research with adults, there appear to be important differences in treatment sensitivity among a number of commonly used methods. For example, in the treatment of agoraphobia, client ratings of anxiety suggest much greater clinical change than do behavioral avoidance tests, and, in the treatment of depression, client self-report measures typically show greater evidence of intervention effects than do clinician ratings (Hill & Lambert, 2003). Further work is needed with child disorders, but these types of discrepancies emphasize the need to match the assessment methods to the purpose of the assessment.

Disorder or Problem Specificity

In line with the foregoing statements, and consistent with the stance taken with EBTs, we see EBAs as being disorder- or problem-specific (however, see our following discussion of caveats), while also recognizing that most childhood disorders do not present in neat packages and that comorbidity is the rule rather than the exception (Jensen, 2003; Youngstrom, Findling, & Calabrese, 2003). A problem-specific approach is consistent with how most assessment and treatment research is conducted and would facilitate the integration of EBA into EBTs (cf. Kazdin & Weiss, 2003; Mash & Barkley, in press). Indeed it is becoming increasingly common to see problem-specific assessment methods advocated in the child treatment literature. Evidence of psychometric adequacy is an extremely important consideration in the development of EBA guidelines but so
too is the evidence base from the child psychopathology and treatment literatures. For an assessment technique to have value in the context of EBA, it must assess a construct shown in the empirical literature to have relevance to the diagnosis or screening of the disorder, case conceptualization and treatment planning, treatment monitoring, prevention, or evaluation of the outcome of treatment. In other words, the assessment technique must provide information crucial to the delivery of services to children with a particular disorder.

Although formal diagnostic systems (such as the Diagnostic and Statistical Manual of Mental Disorders (4th ed., text rev.; American Psychiatric Association, 2000a) provide one alternative for framing the range of disorders and problems to be considered, there is no need to limit the range of problems to those detailed in a diagnostic system. Refraining from excessive reliance on formal diagnostic systems is warranted given the well-documented shortcomings in the nature and development of such systems (e.g., Beutler & Malik, 2002; Mash & Dozois, 2003; Scotti, Morris, McNeil, & Hawkins, 1996) and the lack of evidence that such diagnostic systems provide the best way to match a treatment to a child (Bickman, 2002). Thus, subclinical conditions—for example, adolescents who display symptoms of depression but fail to meet the diagnostic criteria for major depressive disorder (Fergusson, Horwood, Ritter, & Beautrais, 2005; Gotlib, Lewinsohn, & Seeley, 1995) commonly experience emotional and relational problems such as excessive anger, loneliness, conflictual relationships, and problems that do not fit neatly into existing diagnostic categories such as temperamental irritability (Evans, Heriot, & Friedman, 2002)—may also be the focus of EBAs.

There are, however, three caveats to our call for problem specificity with respect to EBAs. First, an important purpose of many assessments is to identify the precise nature of the problem(s). Thus, it is necessary to conceptualize multiple, interdependent stages in the assessment process, with each iteration of the process becoming less general in nature and more focused on elements hypothesized to be central to the client functioning and concerns (Mash & Terdal, 1997a). Accordingly, guidelines for initial assessments may be referral based and relatively problem nonspecific but would become increasingly problem specific with further assessment (Hawkins, 1979; Mash & Terdal, 1997a). Thus, some assessments for diagnostic purposes are likely to begin with very broad, general measures such as multidimensional screening tools, child behavior inventories, or structured interviews. However, even broadband measures must demonstrate appropriate forms and levels of reliability and validity for the various facets or constructs encompassed in the instrument.

Second, children and adolescents who may not meet diagnostic criteria for a well-defined disorder but who have symptoms associated with psychosocial impairment may also experience significant problems (Angold, Costello, Farmer, Burns, & Erkanli, 1999). In some cases, impairments in functioning may be better predictors of adverse outcomes (e.g., school failure) than a youth’s formal diagnosis (Vander Stoep, Weiss, McKnight, Beresford, & Cohen, 2002). Thus, assessments of children and adolescents need to focus not only on specific disorders and problems but also on specific impairments that may occur in the absence of a diagnosable disorder.

The third caveat is that, for some generic assessment strategies, there may be research to indicate that the strategy is evidence based without being problem specific. Illustrative examples, for diagnostic assessment purposes, may include strategies such as functional analysis (e.g., Haynes, Leisen, & Blaine, 1997; Scotti et al., 1996) or case formulation (e.g., Cone-Roberts, Smith & Weisz, 2003; Persons & Bertagnolli, 1999). For treatment monitoring and evaluation purposes, there are a number of recently developed patient monitoring systems that might now or in the near future meet evidence-based criteria (e.g., Borkovec, 2004; Lambert, 2001).

The Scope of the Assessment

A narrow focus on assessing symptoms and symptom reduction is insufficient for both treatment planning and treatment evaluation purposes (cf. Kazdin, 2003a, 2003b). To ensure that the assessments of children, for whatever purposes, are comprehensive and clinically useful, it is necessary to achieve a sufficient understanding of the child and the child’s life context. If guidelines are developed for disorder-specific or problem-specific EBA strategies, it then becomes possible to draw on the relevant developmental and developmental psychopathology literatures to obtain a fuller picture of what the assessment should entail. For example, in the case of a child with attention deficit hyperactivity disorder, this could include indications of key symptoms to assess (e.g., inattention, hyperactivity, impulsivity), as well as common comorbid diagnoses (e.g., conduct disorder, learning disorder) and associated features (e.g., motor incoordination, social problems) that should be examined. Additionally, it could include indications and methods for assessing broader life-context factors consistently found to be relevant to the functioning of children with attention deficit hyperactivity disorder, such as overall quality of life, life stress, family functioning, peer relations, academic functioning, or health care utilization. Unfortunately, many diagnostic systems and conventional tests, particularly those for young children, provide minimal information about the child’s broader life context and have no evidence base in relation to intervention (Neisworth & Bagnato, 2004).
EBAs also need to be sensitive to key characteristics of the child, including age, gender, ethnicity, and culture (Bell, Foster, & Mash, 2005; O’Donohue, 2005; Sonderegger & Barrett, 2004). The developmental psychopathology literature could provide valuable guidance on how these characteristics might influence the scope of the assessment and the translation of the assessment data into a workable treatment plan. For example, findings that girls are more likely to use indirect and relational forms of aggression such as verbal insults or gossip than physical aggression (Crick & Nelson, 2002) may suggest different assessment targets and treatment goals for girls versus boys with conduct problems. At a minimum, attention to gender and other characteristics requires that there is evidence that the assessment methods possess adequate psychometric properties when these characteristics are considered. Attention would also need to be paid to the possibility that a method may not have comparable predictive validity for different demographic or ethnic groups. Unfortunately, the extent of these types of evidence is far from ideal for most clinical methods, as even the most commonly used intelligence and personality tests have tended to underrepresent ethnic minorities in their validation studies (Okazaki & Sue, 2000).

Obtaining assessment data on the kinds of factors described here would allow for a comprehensive evaluation that could guide clinical service decisions, serve as a baseline for determining treatment effectiveness, and possibly provide guidance on intermediate goals in treatment. However, the desire for a comprehensive evaluation must be balanced with clinical realities, such as time constraints in gathering assessment information before, during, and after treatment. The trend toward utilization of brief, face-valid measures reduces this burden somewhat, but there remains a need to prioritize the variables on which assessment data should be collected. Further details on these and related matters can be found in the articles in this special section. Unfortunately, the current limited state of the literature on the incremental validity of assessment with children can provide little guidance to the clinician on best assessment practices or on how to minimize unnecessary redundancies in assessment data (Johnston & Murray, 2003).

In lieu of this, it still may be possible to obtain some evidence-based guidance on what should be considered primary and secondary (or ancillary) assessment targets for a given disorder or problem; that is, what variables must be assessed to ensure an adequate evaluation for diagnosis, treatment planning, treatment monitoring, or treatment evaluation and what variables may be less critical to assess but potentially relevant in a given clinical situation. The child psychopathology and treatment literatures provide abundant sources of information on such issues; therefore, it should be possible to develop appropriate assessment guidelines for clinicians, perhaps based on expert panel consensus statements. For example, the literature on treatments for adolescent depression may provide indications on whether monitoring mood, negative cognitions, self-esteem, or stressful life events should be considered primary or secondary assessment targets (Rudolph, Hammen, & Daley, in press). Likewise, the literatures on family relationships and the treatment of family conflict may yield directions for whether attachment style, the frequency with which withdrawal is used as a relational coping strategy, or the frequency of negative attributions for partner behavior should be routinely or selectively evaluated and monitored (Mash & Johnston, 1996; Snyder, Heyman, & Haynes, in press). The articles in this special section provide guidance on which constructs are most relevant to the assessment of a given disorder.

Assessment Process Parameters

To be truly useful to clinicians and to ensure that, based on the assessment literature, the amount and timing of assessment activities is optimal, EBA guidelines should provide information on issues such as the duration of any baseline measurement, the frequency of assessments, and the duration of any planned follow-up assessment(s). Decision rules for the frequency of assessment and the duration of follow-up are particularly important in child assessment, as studies have found that many children display chronic problems and as many as two thirds of children may show deterioration or relapse 5 years after treatment (Bickman, 2002). Thus, the extent to which measures can be used repeatedly and with children of different ages is an important consideration in the context of monitoring progress during treatment as well as long-term outcomes following treatment.

Little is known about the dose–response relation in assessment (i.e., correlation between the amount of assessment a child receives and the child outcome), although it is often assumed that more assessment is better than less. However, studies of the dose–response relation in child treatment suggest that there is not necessarily a relation between the amount of treatment provided and the outcome (Bickman, Andrade, & Lambert, 2002), and the same may be true for assessment. Not only is there an absence of data regarding the incremental validity of using additional measures or using the same measure repeatedly, it is possible that the accumulation of greater amounts of information in the clinical context may serve to reduce accuracy while increasing judgmental confidence (Nisbett, Zukier, & Lemley, 1981).

Although parameters such as the amount and frequency of assessment may have little impact on treatment outcome in some circumstances, they do address important cost elements (both monetary and opportu-
nity costs) that are highly relevant to clinicians. The repeated administration of a copyrighted symptom measure adds to the cost of the treatment and may, if administered during the treatment session, take time that could otherwise be devoted to other intervention activities. On the other hand, repeated assessment is necessary to gauge the unfolding impact of treatment, to determine whether intermediate goals have been attained at both the individual and system levels (Bickman, 2002; Kazdin, 1993), and to identify factors that may be mediating the effects of treatment outcome. Ongoing assessment also allows for the possibility of the early identification of failing treatment and implementation of appropriate remedial actions (Mash & Hunsley, 1993); indeed, there is growing evidence in work with adults that, if provided with feedback that treatment is proceeding in a suboptimal manner, clinicians are able to make changes that enhance the effectiveness of their treatments (e.g., Lambert et al., 2003).

Other parameters could also be addressed for each childhood disorder or problem, including the use of multiple perspectives (e.g., child, parent, teacher, therapist, and so on) and the appropriateness of exclusive reliance on self-report when psychophysiological, biochemical, or analogue observational data may be required, when data from multiple settings would be beneficial, and when corroborating data may be valuable or unnecessary. Both the availability and the costs associated with specialized laboratory assessments are additional factors that will influence clinicians’ decisions to include these assessments in their services to clients. Ultimately, although these factors must be considered in assessment-related clinical decisions, the task of selecting appropriate measures should depend on an informed consideration of both the source of the information and the assessment method (Kazdin, 1993).

Possible Cross-Cutting Issues

There are many assessment constructs that, based on evidence in the psychotherapy and mental health services literature, may be relevant for treatment planning, monitoring, and evaluation purposes regardless of client disorder or problem or type of treatment (Foster & Mash, 1999). The following three examples of acceptability of treatment, client satisfaction, and the therapeutic relationship illustrate this point.

The acceptability of treatment, including views of effectiveness, reasonableness, and the possibility of negative side effects, has been shown to influence the use of treatments by professionals and clients’ active involvement in treatment (Kazdin, 2003a). Considering and explicitly evaluating the acceptability of the proposed treatment, to clients and significant others, might be relevant to all interventions. A second example from the realm of social validity indexes is the evaluation of clients’ satisfaction with assessment and treatment services (McMahon & Forehand, 1983). Process and outcome measures of client satisfaction may provide information of relevance to numerous service facets such as adherence to treatment directives, the clients’ decisions to unilaterally terminate services, and the overall impact of the assessment, treatment, and the service setting. However, because treatments may have high client satisfaction but poor outcomes (Pekarik & Guidry, 1999), satisfaction will likely need to be assessed in conjunction with other indexes of outcome.

It is of interest to note that although acceptability and client satisfaction have received some attention in relation to child treatment, these dimensions are rarely if ever considered for assessment, although clearly some parents, teachers, and children may display negative reactions to certain tests and to the forms and questionnaires they are asked to complete during assessment. To a certain extent, collaborative models of child treatment (e.g., Silverman & Kurtines, 1996) are far more common than collaborative models of child and family assessment. In this regard, EBA may help to create a dialogue based on evidence for using or not using certain assessment methods, making decisions about what additional assessments are needed, using assessment information in making treatment decisions, and evaluating treatment outcomes. Clinically, the use of EBA may facilitate shared decision making with parents, teachers, and youngsters; reduce disagreements; build consensus; and help promote realistic expectations during assessment and treatment (Hamilton, 2004).

Finally, many studies in the adult psychotherapy literature suggest that the frequent evaluation during treatment of the therapeutic relationship has the potential for improving treatment effects and reducing premature termination (Nathan, Stuart, & Dolan, 2000). Initial evaluations seem to indicate that obtaining alliance data from clients may contribute to greater client improvement and lower rates of deterioration (Whipple et al., 2003). In the child treatment literature, information regarding the therapeutic alliance is sparse. There are, however, indications that the connection between therapeutic relationship variables and treatment outcome and satisfaction is complex and moderated by the type of problem addressed in treatment, the nature of the outcome variables assessed, and which therapeutic alliance (i.e., youth or parent) is considered (Hawley & Weiss, 2005; Shirk & Karver, 2003; Shirk & Saiz, 1992).

To complicate matters, a lack of correspondence has been reported between youth with conduct disorders and counselor perceptions of the therapeutic alliance, with the nature of disagreement moderated by the youngsters’ age (Bickman et al., 2004). There also are findings to suggest that a therapist’s use of relationship-enhancing verbalizations (i.e., high rates of facilitative statements and low rates of questioning) early in
treatment may predict which parents will continue in treatment (Harwood & Eyberg, 2004) and that monitoring the early parent–therapist relationship may be important in successful treatment completion. Other studies have found that youth perceptions of the therapeutic relationship tend to remain stable over the course of treatment (Bickman et al., 2004). Further studies such as these will be needed before it is possible to determine when and how often relationship variables need to be assessed in treatment; to understand the meaning of differences among youth, parent, and therapist perceptions of the relationship; and, more generally, to establish the utility of assessing relationship variables in child psychotherapy.

**Psychometric Qualities of Measures**

Psychologists are well aware of the need to use psychometrically sound measures as the foundation for their assessment activities. For our purposes, we focus our discussion on how issues of norms, reliability, and validity could be addressed in EBA. Consistent with our previous comments, it is important that factors such as gender, ethnicity, and age are appropriately addressed when these psychometric concepts are considered.

Clinical assessment typically entails the use of both idiographic and nomothetic measures. With idiographic measures (such as individualized measures of symptoms or goals), it is crucial that the same items and instructions are used across assessment occasions. Without this level of standardization it is impossible to accurately determine changes that may be due to treatment effects. In a similar manner, when using a standardized, nomothetically based measure, it is essential that norms, specific criterion-related cutoff scores, or both are available to aid in the accurate interpretation of a client test score (Standards, 1999). For example, norms can be used to determine the client pre- and posttreatment levels of functioning and to evaluate whether any change in functioning is clinically meaningful (Achenbach, 2001; Hartmann, Roper, & Bradford, 1979; Kendall, Marrs-Garcia, Nath, & Sheldrick, 1999). Selecting the target population(s) for the norms and then ensuring that the norms are adequate can be difficult tasks, and several sets of norms may be required for a measure. One set of norms may be needed to determine the meaning of the obtained score relative to the general population, whereas a different set of norms could be used to compare the score to specific subgroups within the populations (Cicchetti, 1994).

Unfortunately, the Standards (1999) provide no guidance for how large or how representative a normative sample should be. When a measure is intended to provide information on whether the symptoms are of a clinical magnitude, it is clear that the normative sample should be representative of the general population; otherwise, there is a substantial risk that reliance on the norms could result in either overpathologizing or underidentification (Wood et al., 2002). Regardless of the population to which comparisons are to be made, a normative sample must be truly representative of the population with respect to demographics and other important characteristics (Achenbach, 2001). Ideally, whether conducted at the national level or the local level, this should involve probability-sampling efforts in which data are obtained from the majority of contacted respondents. As those familiar with psychological instruments are aware, such a sampling strategy is rarely used for the development of test norms. The reliance on data collected from convenience samples with unknown response rates renders the resultant norms of questionable value. Therefore, at a minimum, it would be invaluable to clinicians if EBA criteria indicated the quality (and likely accuracy) of the norms for a measure.

Reliability refers to the consistency of a person’s score on a measure (Anastasi, 1988), including whether (a) all elements of a measure contribute in a consistent way to the data obtained (internal consistency), (b) similar results would be obtained if the measure was used or scored by another clinician (interrater reliability), or (c) similar results would be obtained if the person completed the measure a second time (retest reliability or test stability). Not all reliability indexes are relevant to all assessment methods and measures, and the size of the indexes may vary based on the samples used. Many psychologists have suggested criteria that could be used to rate the quality of a measure’s reliability indexes. Robinson, Shaver, and Wrightsman (1991), for example, recommended that the internal consistency of a measure (i.e., Cronbach’s alpha) be .80 or above for research purposes.

Nunnally (1978) argued that when considering tests used for classification purposes one should consider the size of the standard error of measurement. For example, with an internal consistency value of .90, the standard error of measurement is approximately one third of a test score standard deviation. Thus, for decisions that are made on the basis of specific test scores, Nunnally recommended that a reliability of .90 should be the minimum acceptable level. In a thorough review of clinical, personality, neuropsychological, intelligence, educational, and related measures, Charter (2003) found that the median for internal consistency values was .85. If Nunnally’s .90 standard was used, almost three fourths of measures would have suboptimal internal consistency values. Whatever reliability criteria are ultimately adopted, as we stated previously, EBA criteria must be sensitive to the purpose of the assessment and, accordingly, EBA guidelines should indicate how reliable a measure is likely to be when used for various purposes.
In their discussion of validity, Foster and Cone (1995) drew an important distinction between representational validity (i.e., whether a measure really assesses what it purports to measure) and elaborative validity (i.e., whether the measure has any utility for measuring the construct). Assuming that representational validity has been established, it is elaborative validity that is central to clinicians' use of a measure. Accordingly, replicated evidence for a measure's concurrent, predictive, discriminative, and (ideally) incremental validity should be available to qualify a measure for consideration as evidence based. Although demanding, this provides some guarantee of relevance and utility and would help the clinician to attend to the most established measures among the plethora of options available in the literature. We have indicated repeatedly that validation is a context-sensitive concept—inattention to this fact can lead to inappropriate generalizations being made about a measure's validity. There should be, therefore, replicated, elaborative validity evidence for each purpose of the measure and for each population or group for which the measure is intended to be used.

For EBA guidelines to be useful to clinical child psychologists, the guidelines will need to provide concise summaries of the empirical evidence for the assessment task under consideration. With respect to detailing a measure's psychometric properties for a specific purpose, it should be possible to develop clear and informative psychometric profiles. Robinson et al. (1991), for example, developed general rating criteria for evaluating attitude measures. Using a 5-point scale of 0 (none), 1 (minimal), 2 (moderate), 3 (extensive), and 4 (exemplary), they evaluated measures on the basis of test development factors, norms, reliability, and validity. The type of psychometric properties described previously in this section could easily be summarized by means of a comparable scale, using explicit and consensually agreed upon criteria for each descriptor on the scale.

Using either tables or figures, it should be possible to provide a clear depiction of the psychometric status of a measure for various assessment purposes and client groups. The primary challenge in undertaking such a task relates to achieving the right level of detail for the end users of the information (as the editors of Consumer Reports have done in providing their annual ratings of new automobiles). For example, using a map analogy, it is clear that if an inappropriate scale is chosen the resulting map can end up providing too much resolution or too little; if one wishes to find directions to a nearby hospital, neither a national map nor a high-resolution regional contour map would be particularly useful. Although the number of elements that could be included in summary tables or figures could be staggering, it should be possible to strike a balance between overwhelming the clinician with details and boilerplate statements that the measure is psychometrically acceptable. The alternative to undertaking such a summary is, of course, leaving time-pressed clinicians to their own devices in wading through the ever-increasing volume of assessment research.

Figure 1 provides an example of how essential psychometric information (for a hypothetical measure

![Figure 1. Example evidence-based assessment profile for a hypothetical measure of depressive symptoms.](image-url)
of depressive symptoms) could be presented in such a format. Essential psychometric features are listed on the y-axis for studied groups from a general-population sample of adolescents, depressed adolescents in outpatient psychology clinic samples, and depressed adolescents in inpatient psychiatry samples. Because of the limited empirical evidence for assessment processes, we have chosen not to include this important element in our prototype. The level of evidence supporting the measure, shown on the x-axis, could be categorized as 0 (nonexistent), 1 (limited), 2 (adequate), or 3 (well established). The exact meaning of this categorization would depend on the criteria established for the rating of each psychometric element. For example, a rating of 3 for the norms might indicate that a large, nationally representative sample was available, whereas a rating of 3 for internal consistency might indicate replicated findings in which the alpha of the scale is above .85. Clear standards defining what is necessary to attain each rating level could be developed to ensure consistency across those making the ratings, much as is currently done with the standardization of procedures for reviews submitted to the Cochrane Collaboration.

The kind of information depicted in the figure, combined with (a) text or tabular information on evidence, across key client characteristics, for assessment process parameters and the full range of assessment purposes, and (b) the type of data contained in ROC curves (when indicated) could quickly convey the type of information needed by clinicians in selecting and interpreting clinical measures. Additionally, it could do much to enhance clinicians' assessment practices and could render the task of keeping updated with the child assessment literature a much easier and manageable task. If this information was available via searchable, menu-driven Web sites or software packages, it would be an easy task for clinicians to rapidly access the needed information for a given assessment. The use of a Web site could also facilitate the process of frequent updating of information, much as is done for the review groups such as the Cochrane Collaboration and the Campbell Collaboration.

There are very little data supporting the treatment utility of psychological assessment with children, even though the need for evaluating the utility of assessment and the requisite research designs for doing so have been clear for many years (Hayes et al., 1987; Mash, 1979). At present there is some supporting evidence only for the treatment utility of functional analytic assessment strategies (Haynes et al., 1997) and for the use of diagnosis-driven EBTs (which constitute a form of pseudotreatment utility; Nelson-Gray, 2003). Despite decades of validity research and frequent clinical use, there is no scientific evidence that assessment results from even the most commonly used psychologi-


cal tests have a meaningful impact on the outcome of psychological services.

With these humbling considerations in mind, we suggest that for several reasons there is still merit in using available data to identify EBTs. First, assessment is essential to all forms of intervention, and using empirical evidence to enhance clinicians' assessment activities holds the promise of improved services. Second, even if clinicians use EBTs in their practices, high-quality assessment data are necessary to design, implement, monitor, and evaluate the interventions. Third, assessment data may assist clinicians in selecting the most appropriate EBT (if more than one exists) for a client and may provide new avenues for intervention if prior treatment efforts, with or without an EBT, have been unsuccessful (Nelson-Gray, 2003).

The distinction between treatment efficacy (i.e., treatment effects in contexts favoring internal validity considerations) and treatment effectiveness (i.e., treatment effects in contexts favoring external validity considerations) is now commonplace in the treatment literature. An analogous distinction can be made between assessment efficacy and effectiveness, which refers to the difference between assessment validity and utility (see also Foster & Cone, 1995, distinction between representational and elaborative validity). Current efforts to expand the scientific basis for psychotherapy by examining treatment effectiveness could be mirrored by research efforts to determine the effectiveness of assessment. Of course, such research entails the possibility that cherished assessment methods that are useful for research purposes may be determined to have little feasibility or utility in applied contexts (Kazdin, 2005; Mash & Foster, 2001). We must, however, move beyond reliance on validity data to guide our assessment activities, and, ultimately, our hope is that the sine qua non for any assessment strategy or measure would be in its utility in helping to bring about clinical change—that is, whether the assessment makes a meaningful difference in the lives of the children and families who are being assessed.

Integrating and Using the Assessment

Data: Considering the Clinician

Thus far we have focused our discussion on aspects of discrete elements of psychological assessment, such as the forms and methods of assessment. It should be obvious, however, that even if each method used in an assessment is evidence based, there is no guarantee that the resulting synthesis of information and conclusions are themselves truly evidence based. The limitations of human judgment, including clinical judgment, are well known and have been repeatedly documented in recent decades (see Garb, 1998, for a review). On the other hand, the increased use of methods and processes
that are evidence based would be an important step toward reducing the impact of many of the biases and heuristics that have been shown to negatively affect clinical judgment. For example, EBA guidelines encouraging an increased reliance on repeated assessment with psychometrically sound measures could potentially reduce the biasing effects of the tendency to overemphasize dispositional factors in clinical judgments (Smith & Dumont, 2002). Similarly, the recognition that client and clinician ratings typically yield different estimates of the effects of treatment could influence clinicians’ decisions about the overall impact of treatment and whether further treatment sessions are warranted.

Psychological assessment involves the collection and integration of multiple forms of data from multiple sources and perspectives. A particularly critical issue in child and adolescent assessment relates to how information obtained from different data sources, which is known to differ across informants, should be weighted and combined in clinical decision making (Achenbach & Rescorla, 2001). One difficulty such discrepancies create relates to determining what problems should be identified for treatment. Hawley and Weisz (2003), for example, found that in more than 75% of cases, informants did not agree on a single target problem. The relative weight given to information from different informants is known to interact with the child’s age and with the specific problem or disorder being assessed. For example, in assessing internalizing problems, child reports may need to be given more weight for certain purposes than parent reports, whereas the reverse may apply in the case of externalizing problems. A number of the articles in this special section provide some initial guidance for combining and integrating information in relation to the specific disorder under discussion.

In developing guidelines for weighting and combining information, future work will need to give greater attention to the ways in which differences among informants affect the identification, assessment, and interpretation of children’s behavior, as well as the relations among informant discrepancies, informant characteristics (e.g., parent or child psychopathology), and the measurement and interpretation of treatment outcomes. Understanding the basis for disagreements will be important, as it may reflect the likelihood that some features of a problem may differ across settings, with certain aspects of a problem occurring across settings and others being situation specific. In developing EBA guidelines, it will also be necessary to consider the different methods that have been used to assess discrepancies among informants (De Los Reyes & Kazdin, 2004), because such differences could result in different assessment decisions, including whether a child receives treatment and the type of treatment received.

Conclusions and Future Directions

In the preceding sections, we presented a rationale for developing EBA guidelines for children and some elements that we believe need to be considered in developing and disseminating these guidelines. From our perspective, an EBA approach should encourage attention to foundational elements that should be the default (minimal) option in assessment. Looked at another way, we are suggesting a kind of normative-based functional analysis of the problem, while acknowledging that additional variables may need to be considered and assessed in individual circumstances and contexts.

There is now a very large body of scientific evidence relevant to assessment and to the professional contexts in which clinical child psychologists conduct assessments. Consumers of health care services now expect evidence-based services, and those responsible for the administration of health care services require evidence to support the value of any service. This presents a clear and appropriate responsibility for clinical child psychologists to be informed by evidence in their provision of health care services, including their assessment activities. In conjunction with recent efforts to describe and promote EBTs, the development of EBA guidelines could do much to assist psychologists in successfully meeting this responsibility.

Although it may seem a daunting task to produce EBA guidelines, the first step is relatively straightforward. By focusing on specific problems and disorders, a guideline of only a few pages in length, available online or in a software package, could concisely describe primary and secondary assessment targets and, by means of simple menu-driven commands, provide a listing of strategies and methods to be considered for purposes of diagnosis or prognosis, treatment design and planning, treatment monitoring, and treatment evaluation. Accompanying tables could then be used to profile the level of supporting psychometric data for the measures, taking into account the need to attend to characteristics such as age, gender, and ethnicity. To increase the likelihood of adoption and use of such guidelines, they would have to be updated frequently and would have to be simple to access and use. Beyond their potential value to clinicians, these profiles could also be invaluable to researchers and tests publishers, for they would provide indications of gaps in the psychometric evidence for a measure or, indeed, where there is a need for a clinically useful measure of a construct. Although far from exhaustive, an enormous amount of assessment-relevant research is now available to influence psychological practice—the challenge facing us is to synthesize this information into a form that can be readily disseminated and then utilized in a range of health care environments.

In developing criteria for EBA it will be important to distinguish between evidence-based methods and
EVIDENCE-BASED ASSESSMENT

EBA processes. Much of our discussion has focused on evidence-based methods, in part because methods and specific measures are more easily identified than processes and because the focus in the assessment literature has been on psychometric properties of methods and specific measures such as reliability and validity. Assessment processes include how measures are used, as well as other processes describing the way in which assessors gather and use information in the service of a particular clinical purpose (in making strategic decisions in relation to treatment). It will be necessary to describe assessment processes in greater detail if the true utility of assessment is to be determined. Before we know whether assessment practices make a difference in treatment outcomes, we will need to have a better understanding of what processes are being used and how they are being used for particular child populations and in particular clinical contexts.

The articles that follow provide a strong foundation for developing evidence-based guidelines for assessment methods and processes in relation to specific child and adolescent problems. We acknowledge that, at this point in time, it is not possible to require that an assessment, in its entirety, must be evidence based; otherwise we are left with few if any possibilities for conducting child assessments. What we are advocating is a disorder- or problem-specific guideline approach that specifies the minimum necessary assessment processes and methods needed to adequately address the goals of assessment, which would differ depending on the assessment purpose (e.g., diagnosis, treatment planning, treatment monitoring, treatment evaluation) and stage of assessment. Building on the growing literature that demonstrates the validity of numerous psychological tests in applied contexts (e.g., Meyer et al., 2001), the focus of this special section is on the development of guidelines for generic assessment strategies that are based on what is known about (a) a disorder's symptoms, etiologies, and comorbid conditions and (b) the individual, interpersonal, and societal resources and contexts that impinge on the treatment of the disorder. The articles that follow focus on the development of integrated evidence-based “assessment strategies” for a range of childhood disorders, giving attention to the psychometric properties of specific tests and measures, common assessment decisions associated with specific disorders, and the utility of assessment for treatment planning, design, and monitoring.

Based on the reactions of many psychologists to the EBT initiative, it is likely that there will be disagreement about some or many of the issues regarding EBA that are raised in this special section. Given the numerous controversies that have surrounded the EBT “movement” (see Chambless & Ollendick, 2001; Rosen & Davison, 2003; Westen, Novotny, & Thompson-Brenner, 2004), we acknowledge that recommendations to begin tackling the issue of EBA may be tantamount to taking on a 900-pound gorilla while still wrestling with a very large alligator. Consequently, we approach this endeavor with recognition and respect for the magnitude and challenges associated with this task and with the expectation that the articles in this special section will likely raise many more issues than they resolve. Our intent in placing these issues on the table is to provide a scientifically grounded starting point for subsequent considerations of EBA with children and adolescents.

References


EVIDENCE-BASED ASSESSMENT


Rosen, G. M., & Davison, G. C. (2003). Psychology should list empirically supported principles of change (ESPs) and not creditable trademarked therapies or other treatment packages. Behavioral Modification, 27, 300–312.


EVIDENCE-BASED ASSESSMENT


Received September 7, 2004
Accepted March 24, 2005

379