

CHAPTER 1

Assessment of Psychopathology in Young Children

MARGARET J. BRIGGS-GOWAN, LEANDRA GODOY, AMY HEBERLE, and ALICE S. CARTER

| | |
|--|----|
| INTRODUCTION | 1 |
| EARLY PROBLEMS MATTER | 2 |
| Progress in Psychiatric Diagnosis in Young Children | 3 |
| IMPORTANT CONSIDERATIONS IN YOUNG CHILD ASSESSMENT | 4 |
| Reliance on Caregivers for Information | 4 |
| Sensitivity to Contextual Influences, Including Caregiving Contexts | 6 |
| DOMAINS OF DEVELOPMENT | 9 |
| SELECTING AN ASSESSMENT APPROACH AND TOOL | 11 |
| Types of Tools | 11 |
| Understanding Psychometric Properties | 12 |
| Reliability | 12 |
| Validity | 13 |
| Validity of Classification | 14 |
| Normatization | 15 |
| Cultural Validity and Cultural Norms | 16 |
| Knowing What Problems Are Really Being Assessed | 16 |
| Response Formats | 17 |
| Summary | 17 |
| ASSESSMENT TOOLS | 17 |
| Screening Methods | 18 |
| Screening Methods Characteristics of Screening Tools | 19 |
| Selected Screening Tools | 20 |
| Comprehensive Dimensional Tools for Assessing Social-Emotional/Behavioral Problems | 23 |
| Selected Dimensional Checklists | 24 |
| Variation in Emphasis of the Domains That Are Assessed | 29 |
| Diagnostic Approaches | 29 |
| Selected Diagnostic Interviews | 30 |
| Psychometric Properties of Diagnostic Interviews | 32 |
| Observational Assessment | 35 |
| Assessing Impairment | 35 |
| CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH | 37 |
| REFERENCES | 38 |

INTRODUCTION

The past 20 years have witnessed a sea change for *young children's* mental health. It is now recognized that early childhood (0–5 years) is a crucial period for the development of self-regulation, a critical set of competencies that have implications for adaptive functioning in school and through the life span. Early childhood is also recognized as a time when psychopathology may begin to emerge and disrupt young children's developmental progress. In addition, enormous progress has been made in demonstrating that, when psychiatric disorders are defined in a manner that is developmentally meaningful, even very young children suffer from psychiatric disorders that are valid, impairing, and clinically very similar to those experienced by older children (Egger & Angold, 2006; Egger et al.,

2006). Indeed, recent research has indicated that psychiatric disorders are just as prevalent in early childhood as they are in school-age children (Egger & Angold, 2006). Moreover, when young children manifest psychopathology that is impairing, it is often persistent and predicts later difficulties once they become of school age. Equally important, there is increasing awareness that these problems can interfere with learning within early childhood and may set in motion a developmental cascade that likely predicts challenges to lifespan functioning in multiple domains. Research focused on specific disorders has driven discovery of neurobiologic substrates, which has further validated the relevance and reality of early life psychopathology (Luby, Belden, Pautsch, Si, & Spitznagel, 2009; Luby, Si, Belden, Tandon, & Spitznagel, 2009; Stalets & Luby, 2006). These advances in our understanding of young child psychopathology are paralleled by, and one might argue largely driven by, an explosion in reliable, valid, developmentally sensitive measures for assessing a full range of self-regulation, social-emotional development in young

¹Color versions of Figure 1.1 are available at <http://onlinelibrary.wiley.com/book/10.1002/9781119125556>

2 Assessment of Psychopathology in Young Children

children. Specifically, over the past 15-plus years, a number of instruments have been developed to assess parent and other caregiver appraisals of social-emotional functioning utilizing both questionnaire and interview methods. There have also been advances in observational tools to assess clinically significant emotional and behavior problems. With greater acceptance and building on advances in measurement, we are poised to evaluate the benefits of a broad range of prevention and intervention efforts and see increasing discovery of biological and environmental influences on young children's mental health.

As the field presses forward to address the mental health needs of young children both efficiently and effectively, success will be optimized by a well-informed approach to assessment that (1) acknowledges *contextual factors*, including the caregiving environments at home and in other settings, such as child care and early education environments, caregiver influences on social-emotional functioning and assessment, recent changes in family structure or contextual stressors, and sociocultural factors; (2) is framed within the context of a child's *functioning in other developmental domains*, such as language, cognition, adaptive functioning, health, and sensory; (3) is tailored to the *goals and purposes of the assessment and evaluation setting* (e.g., pediatric clinic, day care center, mental health clinic, or private practice); (4) utilizes *reliable, valid, developmentally sensitive tools*; and (5) employs an approach to *interpretation* that views the whole child in relation to contextual and developmental factors and evaluates his or her capacities and participation in developmentally appropriate activities and settings (i.e., impairment).

A primary goal of this chapter is to help clinicians and researchers determine the most suitable measures to use from a wide array of parent and other caregiver report, observational, and direct assessment measures that are now available. Rather than trying to offer an exhaustive list of all existing measures of social-emotional functioning and psychopathology appropriate for young children, we highlight some of the most widely employed and promising tools and approaches, including those that reflect advances in *screening*, comprehensive *dimensional* parent- and other caregiver-report instruments, and diagnostic approaches to young child evaluation. These assessment tools can be categorized as follows: (1) parent and other caregiver report instruments that focus on general problem behaviors; (2) parent and other caregiver report instruments that focus on specific problem areas or disorders (e.g., anxiety, disruptive behavior); (3) parent and other caregiver report instruments designed to assess both problem behaviors and competencies; (4) comprehensive diagnostic interviews for parents of young children; and (5) observational tools

and methods. Within the first three categories, measures can be further divided according to whether they are brief tools appropriate for *screening* or longer *checklist* tools or *diagnostic interviews* that provide more detailed information. Finally, we will close the chapter with a discussion of ongoing challenges, future directions, and opportunities in research on and clinical applications with assessment of young child psychopathology. Although many researchers and clinicians continue to express discomfort about pathologizing, or labeling, young children, our focus is on assessment tools that enhance the recognition and detection of early emerging psychopathology to address mental health needs in an effort to minimize adverse developmental cascades. Moreover, we argue optimistically that by labeling systematic behavioral patterns observed within young children (rather than labeling individual children) we create the potential to develop and disseminate guidance regarding appropriate contextual supports and specific behavioral interventions that are tailored to the needs of children with different behavioral profiles; these early prevention and targeted interventions can be designed to support family beliefs, values, and goals and children's developmental progress while minimizing child and family distress.

EARLY PROBLEMS MATTER

There is now a consensus among child clinicians that children as young as 2 years of age can suffer from significant social-emotional and behavior problems, or psychopathology. Prevalence estimates of clinically significant problems in nonreferred samples have ranged considerably, from as low as 7% to as high as 26%, depending on whether problems are defined in terms of meeting criteria for psychiatric diagnosis or by exceeding a clinical cutoff on a checklist measure (Briggs-Gowan, Carter, Skuban, & Horwitz, 2001; Egger & Angold, 2006; Gleason et al., 2011; Karabekiroglu et al., 2013; Keenan et al., 1997; Lavigne, Lebailly, Hopkins, Gouze, & Binns, 2009; Wichstrom et al., 2012). Rates also tend to be higher among young children exposed to poverty and other psychosocial risk factors (McCue Horwitz et al., 2012; Qi & Kaiser, 2003; Weitzman, Edmonds, Davagnino, & Briggs-Gowan, 2014). Early social-emotional and behavioral problems are linked with impairment in child and family functioning as well as increased parenting stress and worry (Briggs-Gowan & Carter, 2008a; Briggs-Gowan, Carter, Bosson-Heenan, Guyer, & Horwitz, 2006; Briggs-Gowan et al., 2001; Egger & Angold, 2006; Fuchs, Klein, Otto, & von Klitzing, 2013; Keenan et al., 2007; Lavigne et al., 1996; Luby, Belden, Pautsch, et al., 2009). Moreover, social-emotional/behavior

problems in young children have been associated with concomitant delays in child social-emotional competence (Briggs-Gowan & Carter, 2008a; Briggs-Gowan et al., 2001) and shown to predict poorer social competence in elementary school (Briggs-Gowan, Carter, & Ford, 2011). Intervening in preschool to address both social emotional problems and competencies is associated with greater improvements in both areas, including on experimental tasks of emotion knowledge and social problem-solving strategies (Ştefan & Miclea, 2013). Furthermore, contrasting with the historical belief that young children's difficult behavior is just a phase (Keenan & Wakschlag, 2000), there is consistent evidence that for some children these early emergent social-emotional and behavioral problems are persistent and predict poorer functioning at later ages (Briggs-Gowan & Carter, 2008b; Briggs-Gowan et al., 2006; Kim-Cohen et al., 2005; Lavigne et al., 1998; Mathiesen & Sanson, 2000; O'Neill, Schneiderman, Rajendran, Marks, & Halperin, 2014; Shaw, Lacourse, & Nagin, 2005; Speltz, McClellan, DeKlyen, & Jones, 1999; Spence, Najman, Bor, O'Callaghan, & Williams, 2002; Stalets & Luby, 2006; Wakschlag, Briggs-Gowan, et al., 2008). Notably, persistence has been documented for a wide range of problems, including anxiety, depression, attention-deficit hyperactivity, and disruptive behaviors. Thus, consistent and convincing evidence indicates that young children can and do suffer from a wide spectrum of social-emotional and behavioral problems that are often persistent—the presence of impairment further underscores the importance of early identification and prevention in this developmental period.

Progress in Psychiatric Diagnosis in Young Children

Empirical and conceptual work in the areas of posttraumatic stress disorder (PTSD), depression, and disruptive behavior disorders illustrates the role that developmental factors can play in how psychopathology manifests and how several groups have endeavored to establish the relevance of these psychopathologies in young children.

PTSD

Considerable work by Michael Scheeringa and colleagues documents that children as young as 9 months of age can and do suffer from PTSD (Scheeringa, 2007, 2008; Scheeringa, Myers, Putnam, & Zeanah, 2012; Scheeringa, Peebles, Cook, & Zeanah, 2001). This work highlights the importance of considering the developmental capacities of young children when determining the appropriateness of criteria employed for older children and illustrates the utility of adopting a multi-informant, multimethod approach to assessment (Hunsley & Mash, 2007). Scheeringa et al.'s work in this area has driven important recognition of

developmental factors that affect how PTSD presents in young children. For example, many avoidance and numbing symptoms are either developmentally implausible (e.g., sense of a foreshortened future) or internal in quality (e.g., avoidance of internal thoughts, feelings, or reminders of the event), making them very difficult to identify in young children who have limited verbal skills (Scheeringa, 2008). Scheeringa also noted that these types of symptoms may manifest differently in young children. For example, “markedly diminished interest in significant activities” is often observed as constriction of play, and “feeling of detachment or estrangement from others” is often observed as social withdrawal.

Scheeringa and colleagues further documented the central role that parental reactions play in the emergence, promotion, and maintenance of symptoms in young children (Scheeringa & Zeanah, 2001), highlighting the importance of assessing young children's possible PTSD symptoms in the context of parent– or caregiver–child relationships.

Depression

Luby and colleagues' work has established the presence of early manifestations of clinically significant signs and symptoms of depressive disorders in young children (Luby, Belden, Pausch, et al., 2009; Luby & Navsaria, 2010; Luby, Si, et al., 2009; Stalets & Luby, 2006). Their research addresses the complicated developmental question of whether young children's emotional repertoire is itself sufficiently differentiated to encompass true depressive or elated affect and how to differentiate atypical from normative developmental manifestations. They have further noted that greater variability in young children's mood states calls into question the relevance of duration criteria employed by diagnostic systems developed for older children (Gaffrey, Belden, & Luby, 2011). Luby and colleagues utilized a multimethod, multi-informant approach that included (1) parent reports on both dimensional ratings scales and in an age-appropriate *diagnostic interview*; (2) comprehensive observation of the young child's affective range via a laboratory based temperament assessment, thematic play, and parent–child interaction across structured and unstructured conditions; (3) a developmentally sensitive direct interview for preschoolers; (4) cognitive assessment of the child; and (5) neurocognitive assessment of the child. By employing these techniques, they were able to identify a group of young preschool-age children who met modified depression criteria developed for young children.

Further, Luby and colleagues' empirical data have provided crucial insights into how depression manifests in young children (Luby, Belden, Sullivan, et al., 2009;

4 Assessment of Psychopathology in Young Children

Luby, Belden, Pausch, et al., 2009; Luby, Si, et al., 2009). For example, their work has shown that withdrawal and vegetative symptoms are less consistently evident in young children with Major Depressive Disorder relative to older children. They have also shown that sadness, irritability, and thoughts of death, while evident in young children with depression, also are often present in children with anxiety disorders and attention-deficit/hyperactivity disorder (ADHD). In contrast, anhedonia, feelings of guilt, and psychomotor agitation appear to be fairly characteristic of young children with depression. Changes in activity, appetite, and sleep also have been noted. Anhedonia appears to be a very specific marker of depression in young children: it is commonly present in young children with depression and rarely seen in other psychiatric groups or typically developing children (Luby et al., 2002). A subtype of preschool depression characterized by anhedonia also appears to be particularly severe (Luby, Belden, Pausch, et al., 2009). The work of Luby and her colleagues is notable for inclusion of biological correlates, such as cortisol reactivity and addressing family genetic risk to document the *validity* of the revised young child criteria (Luby, Belden, Pausch, et al., 2009; Luby, Si, et al., 2009; Stalets & Luby, 2006).

Disruptive Behavior Disorders

The study of disruptive behavior disorders highlights the complexity of distinguishing normative from clinically problematic behavior during early childhood when some “problem” behaviors are normative (Wakschlag et al., 2007; Wakschlag, Tolan, & Leventhal, 2010). Historically, concern was raised about whether disruptive behavior problems could be reliably assessed during this period (Campbell, 1990) and whether it may be premature to diagnose children whose challenging behaviors might be transient and diminish over time. At the same time, disruptive behavior is the most common reason for referral of young children to mental health clinics (Keenan & Wakschlag, 2000; Thomas & Guskin, 2001). Wakschlag and her colleagues have pressed for the refinement of diagnostic approaches in a manner that describes early manifestations of disruptive behavior in a developmentally appropriate and meaningful fashion and distinguishes between clinically significant disruptive behaviors and expectable variation in children’s capacity to regulate emotions and behavior as they consolidate their self-regulatory skills (Gray & Wakschlag, in press; Wakschlag et al., 2007). This work and the work of others has demonstrated that disruptive behavior disorders can be reliably and validly diagnosed and are associated with marked

impairment in preschool children (Egger & Angold, 2006; Egger et al., 2006; Keenan & Wakschlag, 2000, 2002; Keenan et al., 2007; Wakschlag et al., 2007). Through a longitudinal study, this group further demonstrated persistence over time in early disruptive behavior disorders. In one study, 55% of those with disorders at baseline continued to meet criteria for disorder one year later (Wakschlag, Briggs-Gowan, et al., 2008). These rates are strikingly similar to levels of persistence observed in school-age children (Briggs-Gowan et al., 2003).

Summary

As our understanding of specific early emerging disorders evolves, it is quite likely that additional diagnosis-specific instruments will be developed, allowing clinicians and researchers to elicit information about behaviors that assist in making differential diagnoses (i.e., are specific to diagnostic conditions). Clearly, multimethod, multi-informant assessment is integral to understanding young children’s development. Moreover, observations of the parent–child or caregiver–child interactions across several contexts must be considered a critical component of any young child assessment of social-emotional problems or psychopathology.

IMPORTANT CONSIDERATIONS IN YOUNG CHILD ASSESSMENT

Though challenges exist when conducting any mental health assessment, several are unique to or require greater consideration when assessing young children, including the following: (1) young children’s *limited communication skills*, which results in greater reliance on caregivers for information about the child’s functioning; (2) young children’s *sensitivity to contextual influences* on their functioning and the resulting importance of assessing context when working with this population; and (3) the influence of *sociocultural factors*, which affect the meaning of the child’s behaviors to caregivers and other family members, the way that caregivers report on the child’s behavior, caregivers’ level of worry about the child overall and in relation to specific behaviors, caregivers’ decisions to seek or not seek services, and the assessor’s interpretation of findings. Understanding these challenges can enhance the evaluation process.

Reliance on Caregivers for Information

The fact that infants and young children have no or limited verbal abilities and metacognitive capacities makes it

difficult or impossible to directly collect information from them about their thoughts and feelings about presenting complaints, historical experiences and behaviors, and contextual events. Thus, in contrast to older children, youth, and adults, caregivers play a central role in providing information about the young child's behaviors across multiple settings and contexts, including crucial insight into when behaviors may have emerged and changed over time and contextual factors that may have exacerbated or attenuated symptoms. Collecting information from caregivers brings unique considerations to the assessment process when evaluating questionnaire and interview assessment tools.

First, it is important to establish the best informant in the child's life to best speak for him or her about a given topic. This is usually determined by asking who has a parental or caregiving role for the child in relation to the area or domain of inquiry, that is, the person who assumes responsibility for meeting his or her physical and emotional needs. In this chapter, the term *parent* refers to biological, adoptive, or foster parents or guardians or extended family members who care for the child on a regular basis. The term *caregiver* encompasses parents as well as individuals who routinely spend enough time with the child to be knowledgeable about multiple aspects of the child's social-emotional functioning and day-to-day behavior. Thus, caregiver may refer to extended family members and child-care or day-care providers who have cared for the child for at least one month and who care for the child on a regular basis. (Note, however, that some measures may list different criteria for classifying someone as a suitable informant on the child's functioning.) Typically, the person who brings a child to his or her evaluation appointment and who consents to the evaluation is a parental figure and should be included in the evaluation process; however, it is also important to assess others who may function as caregivers—particularly when a child experiences care in multiple contexts. These additional individuals should also be included as informants. Moreover, it is important to invite relevant caregivers to participate in any child evaluation, as young children's behavior is more variable across contexts (Clark, Tluczek, & Gallagher, 2004; Tronick, 1989).

An evaluation of the child's larger family context facilitates the process of identifying caregivers. It also provides important cues regarding factors such as recent stressors or strained dyadic and triadic family relationships that are currently influencing the child's functioning or that may have had a role in shaping the child's functioning over time (Carlson, 1990; Hayden et al., 1998). Information

about the larger family context also enhances the assessor's understanding of a given caregiver's perspective on the child (e.g., how long the caregiver has known the child, who else is present in typical interactions between the caregiver and the child, who supports or influences the caregiver's beliefs about the child), potentially illuminating sources of bias and avenues for intervention. For example, if in the process of assessing the family context it comes to light that a young mother has received criticism about her child's behavior and her own parenting from her own mother, who cares for the child on a regular basis, this information would point to the need for further assessment of the mother–grandmother relationship. This evaluation could focus on the potential utility of intervention to improve communication between and address inconsistencies in the disciplinary styles, beliefs and expectations of the two caregivers. It is important to note that in using the word *family* we refer to any person, whether biologically related to the child or not, who is considered by the child or other adults close to the child to be a family member. The assessor should consider the total makeup of the family group (e.g., how many members), who is considered a family member, how family members define relationships with one another, where family members live in relation to one another, and what roles family members hold from the perspective of the target child (Carlson, 1990). In addition to evaluating the current family context, a history of the family makeup should be obtained, including ruptures in family relationships, new additions to the family, and changes in the way family members are distributed across households.

In the process of assessing the family and caregiving context and history, assessors must be alert to their own biases about what constitutes a family and how families function. Accurate assessment of family functioning will be facilitated by knowledge of cultural norms regarding family relationships in the groups to which a child belongs. In contrast to the popular notion of the two-parent household and caregiving system, many children are cared for in part or whole by relatives such as aunts, uncles, and grandparents (Dressler, 1985; Wilson & Tolson, 1990). In fact, as of 2012, 10% of all American children reside with a grandparent (with or without a parent also in the home); this number is even greater for Black (14%), Latino (12%), and Asian (14%) children (Ellis & Simmons, 2014). Twenty-six percent of children under the age of 5 whose mothers are employed have a grandparent or other relative as their primary child-care provider (Laughlin, 2013). Many more children may have close but informal caregiving relationships with relatives other than their parents.

Thus, many children have nonparental caregivers who should be included in the evaluation process to elicit an accurate assessment of the child's functioning. The utility of including multiple caregivers' reports in the evaluation process is further discussed in the next section.

Sensitivity to Contextual Influences, Including Caregiving Contexts

Young children's behavior is highly sensitive to contextual and relational influences (Clark et al., 2004; Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012; Gray et al., 2012; Tronick, 1989), and it is therefore critical to include an assessment of relevant contextual factors in any evaluation involving a young child. When working with young children, social-emotional and cognitive functioning and impairment should always be assessed within the context of caregiving relationships and settings, with assessors incorporating what they are able to observe and measure about the various aspects of caregiver functioning and setting characteristics into their interpretation of findings from assessment instruments designed to measure an individual's functioning. In many cases, the family history will reveal recent or chronic stressors to which the child may be responding. Where major events have occurred in the child's history—whether these are identified as stressors by the reporter or not—it may be useful to explicitly inquire about the child's functioning before and after the event as well as changes since the event. As an example, moving houses is a stressful—and common—transition for young children (Stoneman, Brody, Churchill, & Winn, 1999). In some cases, the family history may reveal a recent change in the child's caregiving relationships, as in the case of a child who has recently transitioned to the care of a new foster parent, or in the case of a child whose parents have recently moved the family into their own apartment after having lived with the child's grandparents for the majority of her life. In cases such as these, repeat assessment after one or two months may be warranted to assess the child's functioning after a period of adjustment to the new caregiving environment.

Understanding the makeup of a particular child's family is an essential assessment task because young children's development, behavior, and functioning are embedded within their caregiving relationships. While it is also true that older children are influenced by their caregiving contexts, young children may be especially sensitive to contextual influences due to their dependence on others for their basic needs, often limited exposure to outside institutions such as schools, limited relationships with peers,

and their unlikelihood of having developed independent emotion-regulation or coping strategies that they can utilize in the face of environmental stressors (Compas, 1987).

Caregivers provide structure to the physical, emotional, and behavioral aspects of the child's environment; for example, infants learn to regulate their own affect through exchanges with their caregivers and practice with sensitive, engaged caregivers to repair dysynchronous interactions, to reengage after a rupture, to recover behaviorally and physiologically after exposure to a stressor, and to convey information through affective expression (Bridgett, Burt, Laake, & Oddi, 2013; Haley & Stansbury, 2003; Kogan & Carter, 1996; Martinez-Torteya et al., 2014; Tronick, 1989). As another example, in the behavioral domain, children whose caregivers provide encouragement for independent efforts to achieve goals or learn new tasks may develop a sense of competence and intrinsic motivation that facilitates effortful engagement in challenging tasks in settings outside the home (Ryan & Deci, 2000). Relatedly, consistent with Vygotsky's (1978) theory of proximal development, children whose parents engage in scaffolding behavior (including teaching, praise, encouragement, and positive affective engagement) to enable them to perform within their zone of proximal development—the space between what they have already mastered developmentally and what they are capable of doing with parental support—show superior academic and social self-regulation abilities compared with children whose parents' style of instruction is less contingent on their current level of achievement (Clarke, Kelleher, Clancy, & Cannon, 2012; Neitzel & Stright, 2003).

Caregivers' structuring of the physical environment also may be influential; as an example, low-income children on average have less access to cognitively stimulating books and games (Evans, 2004), which may both impact their actual academic attainment and deflate their scores on assessments due to a lack of familiarity with the types of tasks commonly used in intellectual and academic assessment of young children. A striking percentage of low-income children are in child-care settings that provide inadequate (24%) or minimally adequate (36%) support for their development; in turn, these children show higher levels of internalizing and externalizing behavior problems and demonstrate lower levels of positive behavior compared with their peers in higher quality child-care settings (Votruba-Drzal, Coley, & Chase-Lansdale, 2004). Within all of the areas in which caregivers provide structure for their children, the impact of a given factor is dependent on the match or mismatch between the two; children's temperamental and developmental characteristics may strongly

influence how they respond to a situation. Similarly, caregivers' individual characteristics, including expectations regarding child development and availability of personal psychological resources with which to meet the children's needs for whom they care, will influence whether a set of behaviors is viewed as problematic (Seifer, 2000).

Characteristics of the setting (home, day care, school, and any other place the child spends substantial amounts of time), including both the overall quality of the setting and the match between the child and the setting, may also impact the child's functioning. As noted already, young children's behavior is highly sensitive to contextual and relational influences (Clark et al., 2004; Tronick, 1989), and it may therefore be more variable across settings than older children's behavior. It is common for parents and teachers, for example, to provide very different perspectives, with often very low correlations between their reports (Achenbach, McConaughy, & Howell, 1987). Though this may be due to differences in caregiver behavior or attitudes, it may also, in many cases, be due to differences in the setting, including the physical structure of the setting, the way time and activity is structured in the setting, adult-to-child ratio, number of children, behavior of other children in the setting, level of sensory stimulation, and adequacy of sleep support. Inconsistency in rules between home and school or day care (e.g., requirements for where eating and sleeping take place) may also be a source of differences between caregivers' reports. Given children's sensitivity to contextual influences such as these, the evaluation process would ideally include information gathering not only from multiple caregivers but also across multiple settings and within multiple types of interactions (e.g., unstructured play, challenging goal-directed tasks, shared reading). However, as this goal is often impractical, a compromise is to ensure that data are collected from multiple informants who are familiar with the child's behavior in different settings and in different types of interactions. When gathering information as part of an evaluation, reports from multiple caregivers are likely to yield a more complete picture of the child's functioning than reports from the child's parents alone (Dirks et al., 2012).

It is especially important to gather information from multiple caregivers in cases where a child regularly spends time in different contexts, with different primary caregivers in each (e.g., a child who is in full-time day care with a neighbor, is cared for by his mother at home during the week, and spends weekends in the care of his father). As discussed, where the reports of different caregivers diverge—for example, where a child's father reports problematic levels of disruptive behavior and her

teacher reports no issues with disruptive behavior—the discrepancy should typically be viewed not as evidence of error on the part of one reporter but rather as evidence that contextual, cultural, interpersonal, or other influences may be contributing to the perceived problem. The review of assessment data should then include an analysis of the contexts in which a child is reported to have difficulties, with an eye toward identifying clues for intervention strategies to recommend for the child (e.g., creating consistency between home and day care with respect to the child's nap schedule). In the case in which the child's difficulties occur across multiple caregiving relationships or contexts, they may be more severe than the case in which the difficulties are limited to one relationship or one caregiving context.

Lastly, and briefly, it is also important to consider aspects of the child's context outside of his or her caregiving environment, such as neighborhood and community relationships and functioning. As an example, acts of violence occur considerably more frequently in high-poverty areas compared with more advantaged areas (Briggs-Gowan, Ford, Fraleigh, McCarthy, & Carter, 2010; Friday, 1995; Sampson, Raudenbush, & Earls, 1997), and some studies have reported extremely high rates of community violence exposure among children in high-risk urban areas, with estimates ranging from 42 to 78% (Schechter & Willheim, 2009). Based on these figures, the possibility that a child's behavioral, academic, or emotional difficulties are trauma related should become particularly salient when a child is known to live in a high-poverty/high-risk urban neighborhood; assessors will be even better served by more specific and nuanced knowledge of neighborhood characteristics in the communities in which they work. Even without identifying specific trauma exposure, living in underresourced neighborhoods is associated with elevated disruptive behavior in toddlers (Heberle, Thomas, Wagmiller, Briggs-Gowan, & Carter, 2014).

Sociocultural Factors

In addition to identifying the child's caregiving relationships, mapping out the larger web of family relationships in which the child is developing, and considering the ways the child's presentation may be influenced by the caregiving context and the setting in which the child is cared for, assessors must also consider the ways sociocultural factors may affect caregiver reports, children's performance on assessment instruments, the meaning of a child's behaviors to the caregiver and other family members, whether family members are worried or concerned about the child, family members' service seeking behavior and openness to referrals, and the assessor's own interpretation of the child

TABLE 1.1 Illustrative Examples of Sociocultural Factors that Impact Assessment Findings¹

| | Example of norm | Explanation of relevance in the assessment context |
|---|---|---|
| Communication style | Expectations for appropriate eye contact vary across cultural groups; for example, Asian American children may avoid eye contact with authority figures as a symbol of respect and deference (Thaler, Allen, & Scott, 2014). | Avoidance of eye contact may be interpreted as symptomatic if an assessor from a group in which sustained eye contact is normative (e.g., European Americans) is unaware of the variability in cultural norms for these behaviors. |
| Caregiving norms | Middle-class Puerto Rican mothers have been found to engage primarily in mother-led feeding practices (in which the mother feeds her infant with a spoon or bottle, in contrast to the infant using a spoon or bottle on his or her own) with their 12-month-old infants and to expect that their infants will not yet be able to self-feed at this age (Schulze, Harwood, & Schoelmerich, 2001). | Instruments designed to measure adaptive behavior through the lens of middle-class Euro American culture, in which early attainment of self-feeding skills is emphasized, may falsely identify delays in children whose caregivers have simply not yet organized their interactions with their children to require the development of these skills. |
| Priorities for child behavior | Endorsement of conformity as a child-rearing goal among mothers is negatively correlated with income, educational attainment, and occupational prestige; thus, overall, as socioeconomic status increases, mothers are less likely to see conformity to externally imposed norms as a model for their children's ideal behavior (Luster, Rhoades, & Haas, 1989). | An assessor primed to assess a given caregiver's attitude regarding conformity versus self-direction is likely to have greater success communicating with the caregiver about his or her child's strengths and weaknesses than an assessor who assumes that the caregiver shares his or her own values regarding conformity versus self-direction. |
| Beliefs about mental health services | Latinos, Asian Americans, and African Americans have been found to report high levels of perceived stigma regarding mental health problems in their communities (Alvidrez, 1999; Brown, Marshall, Bower, Woodham, & Waheed, 2014). | An assessor who is familiar with how perceived stigma might function as a barrier to service engagement is likely to be more effective than one who is not at addressing caregivers' concerns about confidentiality and encouraging the family's continued engagement in services (if needed). |

¹Endorsement of cultural norms may vary depending on ethnicity, race, socioeconomic status, immigration status, level of acculturation (for immigrant families), geographic location, and other aspects of identity, both separately and in intersection with one another.

and family's presentation (Table 1.1). Early childhood social-emotional and behavioral problems are identified when a set of behaviors appears with heightened or reduced frequency and intensity or when a child shows developmental deviance through the presence of unusual behaviors or the quality of typical behaviors. While diagnostic guidelines sometimes specify the frequency with which a behavior must occur to be consistent with a clinical diagnosis, it is often the case that caregivers are asked to determine whether a behavior occurs frequently or not, a determination that will necessarily be influenced by the caregivers' expectations for child behavior. Such expectations will vary across cultures and communities. In addition, the determination of whether a behavior (e.g., arguing with adults) has occurred at all is culturally mediated. Thus, the identification of social-emotional and behavioral problems is far from culturally neutral. In addition to problems related to the frequency, intensity, or quality of behavior, children may also show delays in the acquisition of social-emotional competencies (Briggs-Gowan & Carter, 1998). The identification of delays in competencies, like the identification of problem behaviors, is dependent on the caregivers' expectations for the child as well as the child's access to opportunities to develop and demonstrate competencies.

For these reasons, sociocultural considerations must inform the entire assessment process, from instrument selection through interpretation of findings. In selecting instruments for assessment, assessors should consider the following (note that this is not an exhaustive list): (1) Are the items on the instrument culturally relevant for this parent-child dyad? (2) Does the sample on which this instrument was normed represent this parent-child dyad? (3) Is the language in which this instrument is written the optimal language in which to assess this parent-child dyad? (4) If an instrument is designed to use familiar tasks as part of the assessment process (e.g., asking a child to complete visual puzzles), are these tasks actually familiar for this parent-child dyad?

Unfortunately, in many cases socioculturally and linguistically appropriate instruments may not be available. The use of culturally inappropriate instruments may have striking effects and is therefore no small problem; one example comes from a study of Australian Aboriginal children, in which all 124 children who participated in the study screened below the cutoff for detecting developmental disabilities or academic delays on a normed and validated screening tool (the Brigance screen) (D'Aprano, Carapetis, & Andrews, 2011). Such results highlight the damage that may occur when culturally invalid instruments are used

with young children—though, unfortunately, the failure of the instrument to work as intended will likely be far less obvious in the context of one-on-one assessment with a child (as opposed to a large research study or a large-scale universal screening effort). In work with young children and their families, several strategies may be used to enhance the overall cultural appropriateness of an assessment, even if culturally appropriate standardized instruments are not available. For example, it may be helpful to review items with a caregiver, discussing the caregiver's understanding of the meaning of both symptoms that have been endorsed and those that have not. Caregivers can be asked to give examples of a time when they observed a behavior that they have endorsed on a symptom checklist. (Note that this strategy is likely to yield valuable information from parents belonging to a cultural group on which an instrument was normed as well as from parents for whom an instrument may not be culturally valid.) In addition, it is important to ascertain whether the behaviors assessed by a given instrument—whether observed, assessed in an interview, or reported on a questionnaire—are culturally relevant for the target child. For example, if a child is from a cultural group in which nonverbal expression is valued, the examiner should be careful to look for signs that a child is responding nonverbally to a prompt meant to elicit a verbal response in the dominant culture but designed to fit a different cultural group (McLaughlin, Gesi Blanchard, & Osanai, 1995). (A nonverbal response in this case may indicate not a deficit in the skills needed to respond verbally but rather an incongruence between the forms of communication expected by the test designer and the forms of communication familiar to the child.) In addition to attempting to compensate for inaccuracies in a given standardized instrument (or set of instruments) that are not culturally valid for a given client, assessors should also consider whether there is important information about the child's development that is missed entirely by standard measures. For example, standard psychological assessment batteries typically do not include measures of ethnic and racial identity development, although these are essential aspects of the overall development of racial and ethnic minority children (Yasui & Dishion, 2007). Thus, overall, in cases in which the set of available instruments contains no culturally valid tools, assessors must be resourceful in identifying means for gathering supplemental information and extremely cautious about interpreting any standardized results obtained by calibrating a child's functioning against a normative group that does not represent their experience.

In addition to instrument selection, assessors must also consider cultural factors throughout the process of

interviewing parents and children and interpreting the overall findings of the assessment (Kleinman, Eisenberg, & Good, 1978). We suggest that assessors consider the following questions as a starting point for framing a culturally informed evaluation: (1) What does the caregiver see as the problem? (2) What are the caregiver's expectations with regard to treatment? (3) How does the problem affect the child, caregivers, and other family members? (4) What does the caregiver see as optimal behavior for this child? (5) What are the caregiver's beliefs about child development? What does the caregiver see as the key tasks of development at this stage in the child's life? (6) How have community members responded to the child and caregiver? (7) What is the predominant attitude regarding help seeking in the caregiver's community? If known, how do community members see the organization with which the assessor is affiliated? Does the caregiver share these beliefs?

Culturally competent assessment requires that these and related questions be addressed for each family individually; however, familiarity with the cultural groups with which a family identifies is likely to enhance the assessment process by informing the questions asked by the assessor, the way questions are asked, and the way the caregiver's responses to questions are understood. To this end, consultation with colleagues who share an identity status with the caregiver may be invaluable.

In many cases, assessors may have only limited access to information about a child's caregiving context, family relationships, and cultural affiliations. For example, screening instruments may collect only minimal demographic information about a family and are often administered in contexts in which an assessor has little or no direct contact with caregivers. In these cases, the selection of instruments that are appropriate for the target population is particularly important, as is additional, culturally competent assessment of any problems identified by a screening instrument. Care should be taken to ensure that screening results are delivered in a culturally competent manner, with consideration given to the linguistic needs of the family receiving feedback as well as to beliefs known to be widely held in the community that may impact caregivers' understanding of screening results and their actions following the delivery of these results.

DOMAINS OF DEVELOPMENT

One of the myths about early emerging psychopathology that has been debunked in the past two decades of research on young children's mental health is that early emerging

psychopathology is undifferentiated. Indeed, whether validating the presence of differentiation across broad band domains of psychopathology, such as distinctions between disruptive or externalizing problems (which typically include problems in aggression, hyperactivity, inattention, and defiance), mood or internalizing problems (which typically include problems in social withdrawal, depression, and anxiety), and regulatory problems (which typically include sleep, eating, negative emotionality, and sensory symptoms) (Achenbach, 1966; Achenbach, Edelbrock, & Howell, 1987; Carter, Briggs-Gowan, Jones, & Little, 2003) or validating distinctions between different types of symptoms within a dimension of psychopathology, such as anxiety (Mian, Carter, Pine, Wakschlag, & Briggs-Gowan, in press; Mian, Godoy, Briggs-Gowan, & Carter, 2012; Spence, Rapee, McDonald, & Ingram, 2001) or disruptive behavior (Wakschlag et al., 2014; Wakschlag, Henry, et al., 2012; Wakschlag, Briggs-Gowan, et al., 2008; Wakschlag, Hill, et al., 2008), both parent reports and observations support differentiation of psychopathological symptoms in the toddler/preschool period. Based on this relatively new recognition of differentiation in the early emergence of psychopathology (and as reviewed later in this chapter), screening and assessment tools that focus on differentiated aspects of psychopathology have been developed.

Whether used in clinical or research settings, tools that assess social-emotional behaviors and psychopathology need to be interpreted in relation to knowledge about children's relational, family, and cultural contexts and in relation to the children's developmental context, or developmental functioning. We consider the child's developmental context quite broadly, as multiple developmental domains transact with normative social-emotional development and psychopathological trajectories. Risks for or the occurrence of psychopathology intersect with delays, deficits, or atypical functioning across many other developmental domains.

Indeed, intersections in developmental risk across psychopathology and other developmental domains is almost always observed; children with intellectual and developmental disabilities (Einfeld et al., 2006; Emerson, 2003), language delays and specific language disorders (Henrichs et al., 2013; Ross & Weinberg, 2006), learning disabilities (Morgan, Farkas, Tufis, & Sperling, 2008; Yu, Buka, McCormick, Fitzmaurice, & Indurkha, 2006), or sensory processing disorders (Ben-Sasson, Carter, & Briggs-Gowan, 2009; Ben-Sasson, Soto, Heberle, Carter, & Briggs-Gowan, 2014) show increased rates of psychopathology. Importantly, the nature of these intersections varies from one condition to another, so knowledge

of neurocognitive and linguistic profiles or identified intellectual and developmental disabilities can inform which aspects of psychopathology are assessed in greater depth. Thus, if a child has a known neurodevelopmental condition, such as autism or fragile X syndrome, a literature review should help to identify common areas of difficulty in the social-emotional domain (e.g., anxiety) that should be given additional attention in the assessment process. Conversely, if a social-emotional problem such as ADHD is under consideration, problems in executive functioning should be assessed as these two often co-occur.

Direct assessment of multiple developmental domains is often essential to understanding the *whole child*. Further, given that many young children will not be in environments in which informants will have the knowledge to report on the child's functioning across all of the relevant developmental domains, we strongly encourage including direct, norm-referenced assessment of cognitive and linguistic functioning prior to interpreting the results of social-emotional assessments, particularly since some assessments of social-emotional and behavioral functioning assume normative competencies in other developmental domains. Single-informant screeners are not adequate for this purpose. For example, within assessments of attention, questions may include ratings of the child's ability to follow multistep instructions, which may be better explained by a receptive language deficit than inattention. Similarly, a young child with an intellectual disability may not be engaging in pretend play, but this may reflect a general cognitive delay or deficit rather than a specific delay or deficit in the social-emotional domain. In general, we do not expect gains in social-emotional development to exceed those observed in language or cognitive domains. Therefore, these two developmental lines are particularly crucial for contextualizing interpretation of findings in the social-emotional domain.

In clinical settings, assessment of other developmental domains may aid in motivating parents to participate in preventive and targeted interventions designed to reduce psychopathology. In early childhood parents are often more concerned about and more attuned to delays in language development than to atypical emotional, social, or behavioral development (Godoy, Carter, Silver, Dickstein, & Seifer, 2014). Presenting to parents the ways symptoms may be interfering with language (or other) cognitive learning can aid in building a shared view of the child that will motivate parents to participate in interventions focused on reducing symptoms.

Explaining the intersections of developmental lines in relation to establishing risk for psychopathology may also aid in freeing parents from the unfair burden of responsibility or feeling blamed for causing their child’s social-emotional or behavioral problems. For example, explaining the ways challenges in processes such as inhibitory control, executive functioning, language processing, and emotional reactivity can intersect with and heighten risk for psychopathology—particularly given that the explanation can emphasize temperamental and neurocognitive vulnerabilities or adverse environmental exposures—can often facilitate greater empathy for a child who has been engaging in problem behaviors that cause burden to the family system. Therefore, knowledge of multiple developmental domains and their intersections, many of which are well documented in the literature, is extremely important in assessment domain and measurement selection. Moreover, knowledge of children’s functioning in multiple developmental domains is often essential to appropriate interpretation of findings in research and clinical settings, and can be used to promote a stronger alliance between clinicians and parents when giving feedback and offering recommendations about intervention to parents whose children evidence elevated symptoms and disorders in clinical settings.

SELECTING AN ASSESSMENT APPROACH AND TOOL

When selecting tools to assess social-emotional problems in young children, a number of factors should be considered: (1) making sure the tool will best suit the needs and goals and purposes of the assessment and will fit within any setting constraints; (2) reviewing psychometric properties, such as *reliability* and *validity*; (3) evaluating the appropriateness of the tool; and (4) reviewing the *appropriateness of the tool for* children of a particular age or cultural background.

Types of Tools

When selecting a tool, it is important to determine what *type* of tool will provide the level of information required given the goals of the assessment. We discuss three types of tools for collecting information from parents and other caregivers: brief screeners, more comprehensive in-depth checklists, and psychiatric interviews (Table 1.2). These different types of tools vary tremendously in their format, the level of information that they provide, and the purposes for which they are most appropriate. Screeners typically yield scores that indicate whether a child *may* have

TABLE 1.2 General Characteristics of Different Types of Parent and Other Caregiver Report Tools

| Type | General format | Type of information obtained |
|--------------------------|--|--|
| Screeners | 5–10 minutes to complete Completed by parents or other caregivers Minimal training to administer or interpret Many assess a wide array of problems, such as behavior problems and emotional problems Some focus on a specific problem area, such as behavior problems only | Screeners usually provide one or two scores that indicate whether a child falls into a risk category suggesting that the child may have problems without providing detail into the specific nature of the problem. Results often trigger a more in-depth assessment to determine specific areas that are problematic, to identify treatment needs, and to inform planning. |
| Comprehensive checklists | 15–30 minutes to complete Completed by parents or other caregivers Minimal training required to administer Interpretation typically requires training in psychological assessment Provide 2+ broad domain level scores (e.g., externalizing, internalizing) May provide narrowband scores about specific areas (e.g., anxiety) | Comprehensive checklists provide in-depth information about a child’s functioning. Most cover multiple areas, but some focus on specific types of difficulties, such as executive functioning problems. Scoring typically provides a profile of functioning, showing strengths and weaknesses in different areas. These results can be used to establish eligibility for services, generating treatment plans, and documenting change over time. |
| Diagnostic interviews | 1–2 hours to complete at a minimum Parent or other caregiver is interviewed by a trained individual All interviewers must be trained in administration Some may be administered by nonclinicians; others require clinical interviewers Interviews probe for information about whether symptoms are present, duration and onset of symptoms, and impairment | The detailed information obtained from these interviews allows determination of whether a child meets psychiatric criteria for a range of disorders. Scoring also can yield symptom counts within disorders, such as how many depressive symptoms were reported. |

social-emotional problems but do not provide in-depth information about specific areas that are problematic. Longer, comprehensive checklists provide more in-depth assessment of problems within a specific area or areas. Many of these checklists provide detailed profiles that illustrate a child's strengths and weaknesses relative to children of a similar age in multiple areas, such as depression, anxiety, impulsivity, aggression, and social competencies. Such profiles and score details can be especially helpful for treatment planning, and documenting change over time or with treatment. However, checklist tools are *not* appropriate for establishing whether a child meets criteria for a psychiatric disorder. Instead, psychiatric interviews collect very detailed information necessary to establish whether a child meets diagnostic criteria for a range of psychiatric disorders (e.g., symptom presence, onset, duration, and impairment), and can be important for treatment planning. Observational methods also can supplement more traditional informant report methods. In the next section of this chapter, we will review each type of measure in detail and provide a review of selected measures available for assessing social-emotional functioning in young children.

Understanding Psychometric Properties

The first step in reviewing an instrument should be to examine its psychometric properties. Unfortunately, in both clinical and research settings, practical considerations such as a tool's ease of use; the time required for training administration, scoring, and interpretation; costs associated with use; the level of professional training required to administer, score, and interpret the instrument; and the influence of common regional practices often outweigh consideration of psychometric properties. The following section outlines psychometric issues most relevant to parent-caregiver reports in young child assessment: reliability, validity, sensitivity/specificity, positive and negative

predictive value, and *normalization*. Other factors such as cultural considerations, developmental appropriateness, and response formats that may influence instrument adoption also are discussed.

Reliability

Reliability refers broadly to the consistency or stability of a measure. For measures that have scale scores, the consistency of a measure is typically evaluated by examining two or three aspects of reliability: internal consistency, test-retest reliability, and interrater reliability. Table 1.3 provides an overview of these types of reliability and the types of tools to which they generally apply. Rules of thumb for interpreting whether a tool has good enough reliability for different types of measures and related statistics are presented in Table 1.4.

Internal consistency is the extent to which individual items on the scale hang together and reflect the same construct. Usually, this is assessed with Cronbach's alpha (Cronbach, 1951). Scales with Cronbach's alpha of .70 or greater are usually considered to have adequate internal consistency, whereas those with alphas between .60 and .69 are considered marginal and those falling below .60 are considered unacceptable (Cicchetti & Sparrow, 1981; Nunnally, 1978). However, some clinically informative young child measures may not have acceptable internal consistency. This can occur if a measure includes behaviors that rarely occur in the population (e.g., atypical behaviors related to autism spectrum disorder; ASD) or sets of behaviors that are clinically concerning but not likely to co-occur (Achenbach, Edelbrock & Howell, 1987; Briggs-Gowan & Carter, 1998). For example, the ITSEA Atypical Behaviors Index and Maladaptive Index each concern a set of behaviors that are rare but clinically important (e.g., rocking, spinning, PICA, head banging). These indices are expected to have low internal consistency in a normative population (Carter & Briggs-Gowan, 2006),

TABLE 1.3 Types of Reliability

| Type | Definition | Measures to which this is most relevant | Statistic used |
|-------------------------|---|---|---|
| Internal consistency | How well do the items in a scale hang together? | Measures with scale scores, such as screeners or checklists | Cronbach's alpha (α) |
| Test-retest reliability | How consistently does the tool provide the same results over time? | All measures | Correlation coefficients (r) or intraclass correlation coefficient (ICC) (Bartko, 1976) |
| Interrater reliability | How consistently does the tool provide the same results when different individuals rate it? | All measures, especially those where judgment is involved, such as semistructured interviews or observational systems | Intraclass correlations, kappa |

TABLE 1.4 Reliability Statistics and Rules of Thumb for Interpreting Reliability

| Measure type | | Score types | Statistic | Criteria ¹ |
|--------------|---|-------------------------------------|------------------------------|---|
| Dimensional | Dimensional scores | Symptom counts | Intraclass correlation (ICC) | <.40 Poor .40 to .59 Fair .60 to .74 Good .75 to 1.0 Excellent |
| | Test–retest (about the same person at different times) | Coding composite scores | | |
| | Interrater (about the same sample of information about the same person by different raters) | Scale scores | | |
| | Dimensional scores | Items that comprise scale scores | Cronbach’s alpha | <.60 inadequate .60 to .69 poor/marginal .70 Fair .80 Good .90+ Excellent |
| | Internal consistency (how well the items hang together) | Codes the comprise composite scores | | |
| Categorical | Categorical variables | Disorder vs. not | Kappa statistic (K) | <.40 Poor .40 to .59 Fair .60 to .74 Good .75 to 1.0 Excellent |
| | Interrater reliability | Below cut score vs. above cut score | | |
| | Test–retest reliability | Ordinal coding scale | | |

¹Cicchetti, 1994; Landis & Koch, 1977.

despite measuring behaviors that may be red flags for a need for further evaluation and be clinically and diagnostically important in and of themselves. Thus, while internal consistency is one important indicator of reliability, low internal consistency may be tolerated if a score is based on rare behaviors that when present may be clinically informative. Therefore, understanding the items that go into a scale is important to interpreting internal consistency (Cicchetti & Sparrow, 1981).

Test–retest reliability reflects the stability of a measure when it is completed twice over a relatively short time frame. High test-retest reliability suggests that respondents tend to understand items in a similar way over time and that the underlying construct that is measured does not change substantially over a brief period of time. When test–retest reliability is assessed in older children and adults a two- to four-week period is often used. However, a shorter time period is desirable for infants, toddlers, and preschoolers because the rapid pace of development in this period can reduce test–retest reliability. Contextual variation also could contribute to lower reliability between two *assessments*. For example, peer aggression scores may be quite different if a parent answers questions during a time period when the child has not had much interaction with peers (e.g., during summer break) versus when the child routinely attends a child care program. Thus, contextual and developmental variation may contribute to lower reliability over time in young children.

For interview assessments or observational measures, *interrater reliability*, or the consistency of the measure across raters, should also be evaluated. Interrater reliability indicates whether different people who have the same information about a child tend to rate that information

in the same way. In the case of a checklist measure, this might refer to agreement between parents and teachers or between mother and father. In the case of a psychiatric interview, this would mean that after reviewing the parent response, two different people rate the symptom as similarly present or absent. For psychiatric interviews the unit of reliability analysis may be symptom counts or diagnosis. In general, the more structured the assessment format and the lower the inference level required to make ratings, the greater the probability that adequate test–retest and interrater reliability can be obtained in interviews. The availability of detailed administration manuals or training further enhances interrater reliability. A problem often encountered when using clinical interviews is that even the best-trained individuals may drift over time, gradually changing how they administer or rate an interview. This can be minimized by having periodic reviews by an expert who independently reviews and rates the same interview and provides feedback to the interviewer.

Validity

Validity refers to a measure’s ability to measure the problem, skill, or trait that it is intended to measure. Using a measure that does not have adequate validity for measuring a particular problem may lead to mistaken conclusions about the nature of a child’s problems and affect clinical decisions and care. For example, use of a measure with poor validity for assessing anxiety may mean that clinically impairing anxiety does not receive adequate clinical attention, potentially affecting not only anxiety outcomes but also other areas of functioning affected by anxiety, such as participation in age-expected social activities in

which social skills are honed. More accurate, valid characterization of a child's strengths and weaknesses can drive more targeted, tailored treatment planning and optimize outcomes. For research purposes, validity is essential to ensure that constructs of interest are well characterized and optimize capacity to truly test the hypotheses of interest. Several types of validity that are commonly studied in the course of developing social-emotional/behavioral assessment tools are briefly introduced in this section.

Content validity or *face validity* refers to whether the individual items, questions, and scales actually measure the underlying construct they are intended to measure. One aspect of this that is particularly important for young children is whether the items are developmentally appropriate for the target age range. This is often established during a measure's development phase by an independent expert review process. *Construct validity* refers to how well a measure assesses the specific construct it is intended to measure. This can be assessed by comparing scores on one measure with scores on another validated measure of the same problem. Another way to evaluate construct validity is to examine how well a tool classifies children with a particular kind of clinically significant problem or disorder from both other children with other types of clinically significant problems or disorders and from those without clinically significant problems and disorders. This type of validity is evaluated in terms of the measure's sensitivity and specificity, which is discussed in detail later in this section.

Factor structure provides another indicator of validity, by empirically testing whether the structure of the data fits together statistically in a manner that fits with expectations, based on theory or prior evidence in the area. An evaluation of factor structure can test, for example, whether different types of anxious behaviors cluster together into multiple discrete clusters (e.g., with separation anxiety items on one factor and phobic/fearful items on another) that load onto a higher order structure, as has been demonstrated in young children (Mian et al., 2012; Spence et al., 2001). *Convergent validity* refers to whether a measure is positively associated with other measures of related, but different, constructs. For example, given that anxiety and depression symptoms are often correlated, one would expect that a measure of early anxiety would correlate positively with a measure of early depression.

Discriminant validity or *between groups validity* refers to whether a measure differentiates children who are identified as having a particular problem it is intended to measure from children who do not have that problem. For example, one might examine whether children who have been diagnosed with anxiety disorders have higher scores on a

measure of anxiety than healthy children who do not have anxiety disorders.

Predictive validity refers to a measure's ability to predict the same problem over time. For example, evidence that higher anxiety scores on a measure predict higher anxiety later in childhood would support predictive validity. Predictive validity often is not formally evaluated when a tool is being developed, but it can be an important form of validity for young child measures. Demonstrating that a tool predicts later functioning is essential to showing that it captures a clinically meaningful aspect of early functioning that is not simply a transient perturbation in functioning that wanes with development.

Validity Within a Longitudinal Developmental Context

Developmental processes also may affect construct validity. One of the steps in psychometric development is to test and establish the factor structure of a measure, or how items load on scales (e.g., aggression), how scales correlate with one another (e.g., aggression, hyperactivity, oppositional behaviors), and how multiple scales load onto higher order factors (e.g., an externalizing factor that includes aggression, hyperactivity, and oppositionality). However, when following children over time or using a measure that is the same across a broad range of ages, it is also important to evaluate a measure's factor structure over time or between more narrow age groups (Carter et al., 2003). This is rarely reported and it can be difficult or impossible to find information about the consistency of a measure's structure over time. Without knowledge that a measure demonstrates a consistent factor structure across development it is impossible to interpret whether longitudinal patterns represent heterotypic or homotypic stability or instead change.

Validity of Classification

Another important consideration is how well a measure classifies a child as *having a problem* or *not having a problem*. Although a test or measure may be a reliable and valid measure of an underlying construct or behavior, there is still an issue of how useful it might be for a given diagnostic or classification purpose (e.g., meets criteria for oppositional defiant disorder). This is important for tools that yield categorical scores that indicate whether a child's problems are above or below a clinical threshold indicative of problems. In psychological testing, this most commonly applies to screening tools or longer checklist measures that are used to identify children who may have problems and to trigger an action to be taken, such as further evaluation

or referral. The problem of misclassification and its impact on the person being classified is of particular concern.

Sensitivity, specificity, positive predictive value, and negative predictive value are statistics that pertain to classification. *Sensitivity* refers to the proportion of children who fall within a given problem group (e.g., those known to have a disorder according to another criterion) who are successfully classified by the tool in question. If a measure classifies 85 of 100 children with established anxiety disorders as having high anxiety, it has 85% sensitivity. As sensitivity goes up, false negatives go down. *Specificity* refers to the proportion of children who do *not* have a particular problem who are correctly classified by the tool in question. Therefore, a measure that classifies 90 of 100 children who do not have an anxiety disorder as not high in anxiety has 90% specificity. As specificity goes up, false positives go down.

Positive predictive value (PPV) is the proportion of children flagged as positive by a screening tool who truly have that problem according to another criterion. Thus, if 50 of 100 children who are positive according to the tool indeed have problems, PPV is 50%. PPV is correlated with prevalence rates; that is, as prevalence goes up or down, so does PPV, assuming sensitivity and specificity are constant. *Negative predictive value* (NPV) refers to the percentage of children who screen negative that are truly negative according to another criterion.

A general guideline is that assessment tools used for early identification should have minimum sensitivity and specificity of 70% and false positive rates of no greater than 30% of all screened (Cicchetti, Volkmar, Klin, & Showalter, 1995; *Disabilities*, 2001). Though no minimum PPV exists (to our knowledge) for social-emotional assessment tools, tools should successfully identify a subpopulation in which the prevalence of the problem is more common than in the original population (Milgrom, Mendelsohn, & Gemmill, 2011). Low PPV is usually tolerated when failure to detect a problem is associated with a very poor outcome, such as death. For example, a minimum PPV of 10% has been proposed for ovarian cancer screening because the likelihood of death when this form of cancer is not detected is high (Hensley & Spriggs, 2004; Milgrom et al., 2011). As these examples illustrate, different criteria may be appropriate depending on the purpose of the assessment and the risk of harm associated with failure to detect a problem. Generally, sensitivity should be high (and thus false negatives low) for assessment tools designed to aid in specific diagnoses or treatment recommendations (e.g., structured clinical interviews) because of the potential costs to the child and family of misclassifying

a child who *actually has* a specific disorder as *not having* the disorder. However, when the goal is to screen large samples to identify at-risk individuals regardless of their diagnostic status, it can be challenging to find a balance whereby a sufficient proportion of children with problems are detected, while maintaining an acceptable rate of false positives. When scoring thresholds or cut scores are too high, screeners tend to have high specificity and a low rate of false positives, but low sensitivity, meaning that children are more likely to go unidentified. In contrast, when thresholds are too low, sensitivity will be high, but specificity will be low, leading to misclassification and potential flooding of the service system with false positives. False positives also bring some practical costs to the setting conducting the screening and potential emotional cost to families by raising concern or worry unnecessarily.

Finally, one should not assume that a given test will have stable sensitivity and specificity across developmental periods or over time. This is in part because the prevalence of the condition being assessed often varies across periods. A test that has excellent sensitivity, specificity, and positive predictive value with one age group could function poorly in a different age group, particularly if the prevalence of the problem changes. Thus, it is important to review a tool's classification statistics for your target developmental period.

Normatization

The purpose of norms is to provide a means of comparing individuals with their age-based peer group. In many ways, the concept of norms is central to a developmental perspective for assessment. Although traditionally a norm is considered to represent the average test performance of a standardization sample, norms also represent the average, median, or modal performance of a particular developmental level or stage. This is essential at all ages but is especially important in early childhood because children's social-emotional functioning will change with development within this domain and related domains. For example, as illustrated by our work with the ITSEA (Carter et al., 2003) there were age-related changes in some social-emotional problems, from 12 to 35 months, including increases in general anxiety and decreases in separation distress. We and others have also identified robust age differences in children's social-emotional competencies/skills (Carter et al., 2003). For example, on the ITSEA we observed linear increases in 6-month blocks from 12–17 months to 18–23 months to 24–29 months, followed by a social competence leveling from 24–29 and 30–35 months

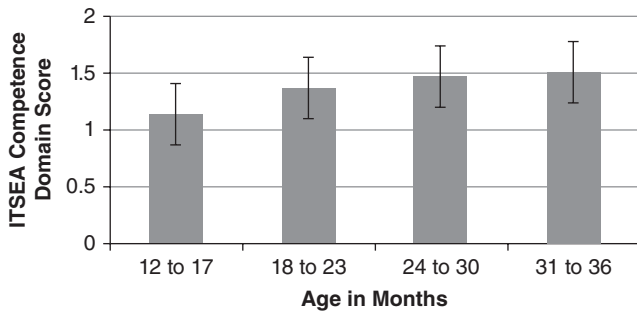


Figure 1.1 Illustration of age effects in social competence from 12 to 36 months of age. See footnote 1.

(Figure 1.1). We also identified sex differences, with greater competence in girls. Thus, a child's competence will be most accurately assessed in relation to children who are the same sex as the child and of similar age within the developmental period from 12 to 36 months.

In selecting a tool and interpreting its results, it is important to be familiar with the sample and conditions in which the normative data were collected—the more similar the normative sample is to your target population, the more relevant the norms (Carter, Marakovitz, & Sparrow, 2006). In a majority of the tests available, the most typical way to operationalize development has been based on chronological age. However, another consideration when comparing a particular child's scores with norms is the child's developmental functioning in other areas. Generally, social-emotional development does not exceed development in other areas. Thus, as noted already, if a child is substantially delayed in multiple areas, such as cognitive and language development, scores may need to be adjusted for developmental level. If a 4-year-old is functioning overall at a mental age of 36 months, then it may make sense to apply 36-month-old norms to determine whether social-emotional functioning is consistent with expectations for mental age. In an ideal world, special norm group data would be available for children with intellectual and developmental disabilities and with genetic and other conditions known to be associated with delays and deficits in social-emotional development and psychopathology.

Normative data also have to be considered within the historical and societal contexts in which they are established. Historical events such as natural disasters, war, or famine may impact the developmental progression of a cohort of children. Norms developed during these periods of relative variation may not be appropriate representations for children growing up in different periods, who are not exposed to these types of experiences. Thus, periodic restandardization of measures is very important.

Cultural Validity and Cultural Norms

Despite a dramatic increase in the diversity of the United States and around the world, issues of cultural validity and the establishment of cultural norms continues to be a challenge to the assessment of social-emotional and behavior problems and disorders in young children. Importantly, translation, even when forward- and back-translation methods are employed, is often inadequate to address issues of cultural appropriateness and validity (cf. van Widenfelt, Treffers, de Beurs, Siebelink, & Koudijs, 2005). Ideally, measures would be developed simultaneously in multiple languages for use in different parts of the world, and examinations of psychometric properties and item and construct equivalence would be established during development. However, there is still extremely limited information about the psychometric properties of measures that have been translated or about the samples or subsamples in which translations have been employed. For example, the ITSEA item "Runs away in public spaces," which is extremely rare in the U.S. context, was rated as normative in Finland until the addition of "for example, in train stations" was added to the item to increase the equivalence of item across these two countries. This is a good example of the limits of translation, as different caregiving patterns and expectations across cultures may lead to situations in which the same words do not have the same meaning in relation to young child psychopathology. Moreover, few measures that have been translated have gathered appropriate normative group data that would permit easy comparisons of the range of relevant psychometric properties across cultural groups. Even when normative data are available for a child and family's home country, these norms may not be appropriate in a new cultural context. For example, for families who have been living in a new country for some time, immigration stress and acculturation may also impact the appropriateness of the use of home country normative data among families. This is clearly an area that will require greater attention in the future.

Knowing What Problems Are Really Being Assessed

A name does not always adequately represent the content of a measure or the scales within it. Therefore, it very important to review the items that make up all of the scales within a measure to truly understand what is being captured and where. In such a review, you may find that a scale labeled "anxiety" includes items such as "cries easily" that could very well reflect other types of problems. Or, an "anxiety" scale might be composed largely or entirely

of items that describe fearful, shy, and socially anxious behaviors but may fail to assess separation anxiety. The term “withdrawal” is often part of a scale’s name—item review is important to determine whether this refers behaviors that appear to reflect withdrawal related to (1) shy, slow-to-warm-up behaviors; (2) depressed, anhedonic behaviors; (3) behaviors that could be related to ASD problems such as lack of eye contact and unawareness of social context; or (4) some or all of these types of difficulties.

For clinical and research applications, another important consideration when goals involve following children over time is whether the tool (or tools) used are measuring the same construct over time. One might use the same tool over time or different tools because a single tool does not cover the whole intended age period. Whichever approach is used, we again recommend reviewing the items that make up the scale at different ages. The items within a scale may change in subtle or dramatic ways even within the same tool.

Response Formats

A final consideration when reviewing a checklist or interview is the type of response format that is used. Behaviors vary in terms of whether they are ever present, how frequently they occur when present, and how intensely they are displayed. Similarly, response formats vary in whether a problem is judged as present versus absent, ratings of frequency, or a judgment of how severely the behavior manifests (Achenbach & Rescorla, 2000; Carter & Briggs-Gowan, 2006). Many measures have a response option that crosses ratings of frequency with ratings of how typical the behavior is of the child. As most social-emotional and behavioral problems and competencies vary in frequency or intensity, a response format that captures variation in frequency and intensity makes some sense (Achenbach & Rescorla, 2000). For example, the response format for the ITSEA is 0 for “Not true/rarely,” 1 for “somewhat true/sometimes,” and 2 for “Very true/often.” A contrasting approach, employed in the Behavior Checklist (BCL; Richman & Graham, 1971) is to ask the parent to rate the degree to which a particular behavior is perceived to be a problem. On the BCL parents are asked to rate a behavioral category as 0 for “No difficulties,” 1 for “Moderate difficulties,” and 2 for “Definitive difficulties.” Although very different, each of these response scales is sufficient for identifying children at elevated risk for psychopathology when items are aggregated into meaningful clusters or scales (Achenbach & Rescorla, 2000; Campbell, 1995). For determining clinical

diagnostic status, however, neither format is sufficient. Rather, information about onset, duration, and intensity of behaviors and impairment are required to determine diagnostic status. Given the complexity of the information that must be gathered to determine diagnostic status, it is usually preferable to gather this type of information in an interview format.

A final consideration is that some response formats can be overly complex, especially for informants with more limited educational backgrounds. Thus, a scale with a response format that changes from one question to the next or utilizes a wide response scale, but with far fewer anchors, can be confusing or off-putting to parents and introduce error variance into measurement. Items with double (or even single) negatives can also complicate a respondent’s ability to use a particular response format (e.g., “Often/Very True” that this behavior is rarely present). Even a simple “Yes/No” format may introduce error, as the threshold for difficulty may not simply be the presence or absence of a behavior but the routine use of the behavior in appropriate contexts (e.g., eye contact). Finally, it is usually easier to rate one attribute or behavior at a time (e.g., child often has angry tantrums) rather than being forced to indicate whether the child is more like one pole of a continuum or another (e.g., whether being angry is more characteristic than being calm).

Summary

Many factors should be carefully reviewed and considered when embarking on a new type of assessment, including the psychometric properties and developmental appropriateness of the tool, how well its norm groups match with the target population (or child). Spending the time to understand the breadth and depth of the information the tool provides and the individual problems that go into different scales or domain scores is essential to a strong assessment foundation.

ASSESSMENT TOOLS

Several types of tools can be used to systematically gather information about social-emotional problems or competencies from parents and caregivers, including brief screening checklists, more comprehensive checklists designed to capture variation both in broad domains and within subareas within those domains, diagnostic interviews, and observational tools designed to capture clinically meaningful information that can supplement parent or teacher reports.

Screening Methods

Screening tools, designed as brief measures for detecting children who may be experiencing problems, are most appropriate when the goal is to quickly and efficiently identify such children from a larger population. Large-scale screening efforts can help to reduce the high rates of unmet mental health needs. Among school-age children, approximately 80% of children needing mental health services are not receiving them. Unmet mental health needs are even greater among young children, children of color, and uninsured children (Kataoka, Zhang, & Wells, 2002). The considerable unmet need for mental health services underscores the importance of enhancing early detection efforts through routine mental health screening.

Universal screening can help detect children at risk for developing mental health problems. Typically, screening yields information about which range a child's score falls within (e.g., normal, borderline, or abnormal). Screening should not be used for diagnosis but rather as a jumping-off point for discussions about mental health and further assessment. Once identified, at-risk children benefit from follow-up that can include more in-depth assessment (e.g., longer measure, clinical interview, clinical observations) to evaluate the extent and nature of the difficulties initially detected. This practice, referred to as multistaged or multistage screening, can serve as a cost-effective and minimally burdensome means for screening large groups of children in a manner that permits characterization of risk status (stage 1) and a more detailed profile of problems and competencies when risk is elevated (stage 2). The decision to proceed to later stages of screening can be based on factors such as score profile, parent's indication of concern, and level of impairment. However, even if multistaged screening is not being implemented, screening ideally involves attention to the child's behaviors, as well as contextual risk factors in the family and community, including exposure to violence, parental psychopathology, and neglectful parenting.

In addition to identifying the existence of concerns, screening can help to facilitate appropriate follow-up, such as referral for treatment, by identifying potential domains of concern (e.g., speech-language versus gross motor or anxiety versus disruptive behaviors) and severity of concerns. Agencies and providers employing multistage screening may wait until more comprehensive assessment is completed to make referrals, but in many cases second-stage screening does not occur and providers choose to use the results of screening to initiate a referral or begin

addressing identified concerns. For example, within pediatric primary care settings, providers may refer a child to the local early intervention program following elevated developmental screening. However, in most cases, the places to which a provider is referring will do a more comprehensive assessment before beginning treatment. While screening tools can help in the treatment planning process and in the documentation of treatment gains, given the brevity of screeners and the limited information provided, it is not ideal to use screeners in this manner. Screening is ideally used at multiple points in time to ensure the continued monitoring of concerns among a large group of children.

Screening can occur universally or it can be targeted. In universal screening, all children in a particular setting or situation (e.g., all children seeing their physician for a well-child visit) would complete screening. Ideally, when universal screening is employed in pediatric practices, every parent (or adolescent child) completes a standardized screening tool at each well-child visit from infancy through adolescence, and identified risks or concerns are closely monitored and addressed as appropriate. In targeted screening, only those children identified via other means (e.g., developmental surveillance in which parents report their concerns about the child's behavior) would be screened.

Universal screening is ideal because it does not rely solely on provider judgment or parent expressions of concern, both of which can be flawed means of identifying issues particularly in early childhood. A review study of screening in pediatric primary care settings determined that sensitivity rates for pediatric providers ranged from 14 to 54%, and that specificity rates ranged from 69 to 100% (Sheldrick, Merchant, & Perrin, 2011). Furthermore, parents of children with mental health problems often do not report concerns or worries about their children's behavior (Godoy et al., 2014; Sheldrick, Neger, & Perrin, 2012). In particular, parents of younger children exhibiting social-emotional/behavioral problems report lower levels of worry or concern than parents of older children and often do not raise concerns with the pediatrician (Ellingson, Briggs-Gowan, Carter, & Horwitz, 2004; Godoy et al., 2014; Horwitz, Gary, Briggs-Gowan, & Carter, 2003; Sheldrick, Neger, et al., 2012). Thus, universal screening can help to identify at-risk children more accurately than when screening is not implemented. Moreover, universal screening can help to reduce disparities in health care because all children receive a similar assessment regardless of family characteristics and provider discretion.

Universal screening is growing increasingly popular, particularly within the pediatric primary care setting.

Pediatric primary care settings are well-suited to the task of detecting social-emotional and behavior problems (“Policy statement—The future of pediatrics: Mental health competencies for pediatric primary care,” 2009) as pediatricians can act as the gatekeepers of mental health for a wide range of children. Pediatricians, who often have longitudinal, trusting relationships with families, regularly discuss social-emotional behavioral issues, and they are often the first professional with whom a parent discusses concerns. Thus, they are the people who can first identify problems and facilitate appropriate prevention and intervention services. Families with children who have social-emotional and behavioral problems tend to use pediatric services more than other families (Gadomski, Wissow, Slade, & Jenkins, 2010) and therefore have more opportunities for contact with pediatricians. Pediatric primary care may be particularly helpful in the identification and prevention and intervention of early childhood mental health problems given the more frequent visit schedule among children under 3 years. Additionally, prior to entering more formal child-care or school settings, pediatricians may be the only professional with whom some families have contact who can help to identify problems and facilitate referrals. Indeed, parents who discuss their children’s social-emotional/behavioral problems with their pediatrician are more likely to obtain mental health services for their children than those who do not discuss these issues (Briggs-Gowan, Horwitz, Schwab-Stone, Leventhal, & Leaf, 2000).

Universal mental health screening is increasingly important given the growing list of topics pediatricians are expected to cover and shorter office visits. These factors make it more challenging for pediatricians to detect concerns on their own as they have less time to discuss behaviors and make clinical observations and as parents have less time to raise concerns. Additionally, implementing universal mental health screening, particularly in non-mental health settings, can help to destigmatize mental health and increase the likelihood that parents will raise and discuss concerns and seek treatment as needed.

Yet universal screening efforts can be hampered by logistical, systemic, and other barriers. Logistical barriers include issues like difficulty integrating screening into an already busy practice workflow. An additional barrier is pediatricians’ self-reported lack of confidence and training in addressing mental health issues (Horwitz et al., 2007). Systemic issues include lack of adequate mental health resources, which is even more of an issue with early childhood mental health.

Despite these challenges, universal mental health screening in primary care has been shown to be feasible and effective in improving rates of referral (King et al., 2010; Schonwald, Huntington, Chan, Risko, & Bridgeman, 2009). Not surprisingly, universal screening is gaining popularity within other early childhood settings, such as child care centers and women, infant, and children’s offices. Given the many contexts in which children may be screened, it is important that states and local agencies work together to ensure that screening efforts are coordinated (e.g., discussing screening across settings, sharing results across agencies through a database or other means).

Screening Methods Characteristics of Screening Tools

Screeners should have several characteristics, especially if they are to be implemented widely. They should be brief and easy to administer, and most parents should be able to complete them independently. They should also be easy and quick to score and interpret. Ensuring these characteristics of a screener can ensure that a maximum number of screens are completed and that staff are not overly burdened. Advances in technology have led to moving screeners to an electronic format, which can further facilitate completion and scoring.

Screening tools vary in the scope of their assessment. Some screening tools are designed to detect a narrowly defined problem. For example, the Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F; Robins et al., 2014) is designed to detect ASD and would not be appropriate when the goal is to detect children with social-emotional or behavioral problems in areas outside of ASD. In contrast, more global screeners, such as the Ages and Stages Questionnaire: Social Emotional (ASQ-SE; Squires, Bricker, & Twombly, 2002) and the Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan & Carter, 2006), are designed to identify children who may be evidencing problems in the broad domain of social-emotional/behavioral problems. Some screening tools, such as the BITSEA and the Strengths and Difficulties Questionnaire (SDQ; Goodman & Goodman, 2009), go a step beyond problems to also detect social-emotional competencies. Selecting a screening instrument that meets the goals of the screening process (narrow versus broad, problems versus competencies) is essential to ensure a successful screening initiative.

In recent years the number of screening tools available for the infant-toddler period has grown. However, many screeners now available focus on a specific area of social-emotional adjustment (e.g., behavior problems) and

do not provide comprehensive coverage of both problems and competencies. For example, the 35-item Eyberg Child Behavior Inventory (ECBI; Eyberg & Pincus, 1999), and the 40-item Toddler Behavior Screening Inventory (TBSI; Mouton-Simien, McCain, & Kelley, 1997) measure behavior problems in children 2 to 16 years of age and 1 to 3 years of age respectively whereas the Preschool Anxiety Scale (PAS; Edwards, Rapee, Kennedy, & Spence, 2010; Spence et al., 2001) measures anxiety symptoms in children 3 to 5 years of age. While these scales all demonstrate acceptable psychometric properties, they may be limited in their ability to screen for a wide range of problems and address social-emotional competencies. Additionally, screening tools that were designed for older children are often scaled down for use with younger children without adequate assurance regarding the developmental appropriateness of items nor the psychometric properties within a younger population.

Selected Screening Tools

Table 1.5 displays several commonly used, brief, parent-report social-emotional/behavioral screening questionnaires. We selected tools that cover an array of problems (and, in several cases, competencies) and that may be appropriate in settings in which wide-scale screening is occurring. These questionnaires are also described briefly. All the questionnaires selected can be completed by parents (and in some cases child care providers) on their own or with the assistance of a staff member. They all usually require 5–10 minutes to complete and are easy to score. Qualified individuals who have undergone appropriate training can score the instruments. Regarding interpretation, all the measures require knowledge of standardized assessment and supervised training in working with relevant clients.

The Ages & Stages Questionnaire–Social-Emotional Version (ASQ-SE; Squires et al., 2002), a screening tool that can be completed by parents of children 3 to 66 months of age, addresses seven behavioral areas: self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interaction with people. There are eight versions of the questionnaire that correspond with different child ages (6, 12, 18, 24, 30, 36, 48, and 60 months) and eight corresponding scoring sheets. The number of items varies depending on the questionnaire ranging from 22 to 36. For each question item, respondents are asked to indicate whether the behavior is something that the child does “Most of the Time,” “Sometimes,” or “Rarely or Never.” In addition, respondents can indicate whether

the behavior described by each item is something that is a “Concern.” Several additional questions at the end of the questionnaire ask about parents’ concerns about child social-emotional/behavioral functioning (e.g., “Do you have any concerns about your child’s eating and sleeping behaviors or about her toilet training?”). An overall summary score can be compared with age-specific cutoff scores to identify children at risk for social or emotional difficulties. The ASQ: SE has strong internal consistency (.82), test–retest reliability (.94), sensitivity (.82), and specificity (.92) in detecting children with developmental delay or social-emotional diagnoses (Squires et al., 2002).

The BASC-2 Behavioral Emotional Screen System (BASC-2 BESS; Kamphaus & Reynolds, 2007) measures behavioral and emotional strengths and weaknesses from preschool through high school. In the early childhood age group (ages 2 to 5 years), tools are available for parents and teachers to complete. The tool asks about internalizing problems, externalizing problems, and adaptive skills, though questions in these domains are combined into a total score. Validity indexes can be used to identify overly negative or inconsistent responses. The BASC-2 BESS has demonstrated acceptable test–retest reliability (0.79–0.85), interrater reliability (0.76–0.90), validity with other measures of child behavior such as the Child Behavior Checklist (0.71–0.76), sensitivity (0.73–0.82), and specificity (0.97) in detecting children with social-emotional/behavioral problems (Kamphaus & Reynolds, 2007).

The Brief-Infant-Toddler Social-Emotional Assessment (BITSEA; Briggs-Gowan & Carter, 2006) is a forty-two-item measure for identifying social-emotional/behavioral problems and delays as well as child competencies in 1- to 3-year-old children. The BITSEA generates a Problem and a Competence score, each of which can be compared with age- and gender-specific cut points. The BITSEA has good psychometric properties, including test–retest reliability and interrater reliability. The BITSEA cut points have demonstrated high sensitivity in detecting children with social-emotional/behavioral problems according to other gold standards (diagnostic interview; Child Behavior Checklist) while maintaining acceptable specificity (Briggs-Gowan et al., 2013). In addition, the BITSEA is highly sensitive to autism spectrum disorders (Carter & Briggs-Gowan, 2005; Gardner et al., 2013). Finally, research suggests that BITSEA has acceptable sensitivity and specificity and predicts a majority (PPV of 67.9%) of children who meet the criteria for a psychiatric disorder, according to the Preschool Age Psychiatric Assessment (PAPA) interview (Briggs-Gowan & Carter, 2008b).

TABLE 1.5 Brief Checklist/Questionnaire Measures That Address Social-Emotional and Behavioral Problems Broadly

| | ASQ-SE Ages and Stages Questionnaire Social-Emotional | BESS Behavioral Emotional Screen System | BITSEA Brief Infant-Toddler Social and Emotional Assessment | ECSA Early Childhood Screening Assessment | SDQ Strengths and Difficulties Questionnaire | SWYC Survey of Wellbeing of Young Children |
|---|--|---|--|---|---|--|
| | Squires, Bricker, & Twombly (2002) | Kamphaus & Reynolds (2007) | Briggs-Gowan & Carter (2006) | Gleason, Dickstein, & Zeanah (2010) | Goodman, 1997 | Perrin & Sheldrick (2013) |
| Age | 3 to 66 months | 36 months+ (18 years) | 12 to 36 months | 18 to 60 months | 24 months+ (18 years) | 1 to 66 months |
| Respondent (for early childhood forms) | Parent | Parent | Parent | Parent | Parent | Parent |
| Available languages ¹ | Spanish | Caregiver/teacher Spanish | Caregiver/teacher Spanish, Chinese, Dutch, French, German, Hebrew, Italian, Russian, Thai | Caregiver/teacher Spanish, Romanian | Caregiver/teacher Spanish and over 40 languages | Spanish, Burmese, Nepali |
| Reading level | 5th to 6th | 4th to 6th | 4th to 6th | 5th | Low (though not specified) | Low (though not specified) |
| Number of items | Approximately 30 items (varies by age) | 25 to 30 | 42 | 36 Teacher 40 Parent | 25 (plus up to eight items about impact of problems on the child and family) | ~40 (varies by age) |
| Time frame covered | Not specified | Not specified | Past month | Not specified | Past 6 months | 2 weeks for PHQ-2 Other areas not specified |
| Item response options | 3-point rating scale: Rarely or Never; Sometimes; Most of the Time | 4-point rating scale: Never; Sometimes; Often; Almost Always | 3-point rating scale: Not True or Rarely; Somewhat True or Sometimes; Very True or Often | 3-point rating scale: Rarely/Not True; Sometimes/Sort-Of; and Almost Always/Very True | 3-point rating scale: Not True; Somewhat True; Certainly True | Variable |
| Normalization age bands | 6- to 12-month age bands | 12 months+ | 6-month age bands | None | 24 months+ | 2 to 12 months |
| Normed for boys and girls separately | N | Y | Y | N | N | N |
| Standardized scores (e.g., <i>T</i> scores) | N | Y | Y | N | N | N |
| Scoring | One score produced, which is compared with clinical cut point that varies by child age | One score produced, which is compared with clinical cut point | Produces a Problem score and a Competence score, each of which can be compared with age- and sex-specific cut points | One score produced for the child-related items and one score for the parent-related items; each is compared to a clinical cut point | Produces five subscale scores (Emotional Symptoms; Conduct Problems; Hyperactivity; Peer Problems; Prosocial Behavior) plus a Total Difficulties score, which represents combination of 4 problem domains | Produces separate scores for each part of the SWYC (developmental milestones, family risk, BPSC, PPSC, POSI) |

(continued)

TABLE 1.5 (Continued)

| | ASQ-SE Ages and Stages Questionnaire Social-Emotional | BESS Behavioral Emotional Screen System | BITSEA Brief Infant-Toddler Social and Emotional Assessment | ECSA Early Childhood Screening Assessment | SDQ Strengths and Difficulties Questionnaire | SWYC Survey of Wellbeing of Young Children |
|--|---|--|---|--|---|---|
| COVERAGE: Problem domains covered | Problems are combined into one total score, though it covers a range of areas including self-regulation, compliance, communication, autonomy, and interaction with others | Problems are combined into one total score, though it covers internalizing and externalizing problems | Problems are combined into one total problem score, though it covers internalizing, externalizing, dysregulation, and maladaptive behaviors | Problems are combined into one total score, though it covers a range of areas | Emotional Symptoms; Conduct Problems; Hyperactivity; Peer Problems | BPSC includes 3 domains (Irritability, Inflexibility, Difficulty with Routines); PPSC covers several domains (internalizing, externalizing, attention problems, parenting challenges) though problems are combined into one overall score |
| Competencies/skills covered | Competencies are assessed but responses are integrated into overall score. There is not a separate competency scale. | Competencies are assessed but responses are integrated into overall score. There is not a separate competency scale. | Includes separate Competence domain with separate cut point | No | Includes separate Prosocial Behavior Scale domain with separate cut point | No |
| Parent worry or concern addressed? | Parent can indicate concern about each item; plus open-ended questions about concerns; not used in scoring | No | Two questions gauge parent worry about child behavior and language functioning; not used in scoring | Parent can indicate concern about each item plus there is one question asking about overall concern regarding child's emotional or behavioral development; not used in scoring | Not explicitly, though Impact Supplement asks whether caregiver thinks child has difficulties in emotions, behavior, concentration, or ability to get along with others | Two questions regarding parent concerns about child's behavior and child's learning or development; not used in scoring though providers are encouraged to follow up on concerns raised |
| Additional notes | Ability for parents to complete forms by listening to audio recordings; includes Validity Index to identify responses that may be overly negative or inconsistent | | Parent report form includes two questions about parent depression (PHQ-2) and two questions about parent stress | Impact supplement can be used to assess distress and impairment associated with symptoms | Charts showing normative curves (percentiles) available for BPSC and developmental milestones available, though PPSC and POSI have one cut point | |

¹ Many measures are available in other languages but are often not normed in those languages.

The Early Childhood Screening Assessment (ECSA; Gleason, Zeanah, & Dickstein, 2010) is a brief screening tool designed to identify emotional and behavioral problems in children 18 to 60 months of age. There are two versions of the questionnaire—a parent report form, which consists of 40 items, and a caregiver/teacher report form, which consists of 36 items. The 36 child-specific items are similar across forms. The parent report form also includes the U.S. Preventive Health Task Force recommended Patient Health Questionnaire—2 questions to assess for parent depression (though the time frame “over the past two weeks” is not included) and two questions that ask about parent stress. Items are scored on a 3-point Likert scale (Rarely/Not True, Sometimes/Sort-Of, and Almost Always/Very True), plus parents can circle the “+” next to each item if they are “concerned about an item and want help with it.” Parents can also indicate at the end of the questionnaire whether they are worried about their child’s overall behavior. Scores on the child questions and the adult questions are separately compared with clinical cut points to determine risk status. Though research on the ECSA is not as abundant as that on other screening tools, preliminary research suggests that the ECSA has adequate sensitivity (86%) and specificity (83%) as well as good internal consistency (0.91) and test–retest reliability (0.81) (Gleason et al., 2010).

The SDQ (Goodman & Goodman, 2009) is a 25-item behavioral screening tool composed of five items in each area: (1) Emotional Symptoms; (2) Conduct Problems; (3) Hyperactivity; (4) Peer Problems; And (5) Prosocial Behavior (strengths). A separate one- to eight-item scale assesses problem presence, level of severity, and impairment (impact supplement). For providers using the SDQ over time when interventions are put in place, a follow-up form is also available, which asks two questions about the use of interventions (whether the intervention reduced problems or helped in other ways). The SDQ can be used with children 24 months to 18 years of age. There are three age groups for which separate forms are available: 2 to 4 years; 5 to 10 years; and 11 to 17 years. Parent report and teacher report forms are available. Scoring can be completed in less than 5 minutes either by hand or online. Results yield scores for the five scales, plus a Total Difficulties score (sum of the four problem domains, excludes prosocial behavior scale), and an Impact score. However, the SDQ authors have reported that in low-risk or general population samples, it may be better to examine two problem scales, internalizing (Emotional Symptoms and Peer Problems) and externalizing (Conduct Problems and Hyperactivity) rather than the four subscales (Goodman,

Lamping, & Ploubidis, 2010). The SDQ has been widely used and investigated cross-culturally with normative data obtained in various countries including the United States using a large sample. While the psychometric properties among of the SDQ among school-age children are strong (Sensitivity = 85%; Specificity = 80%; Goodman, Ford, Corbin, & Meltzer, 2004), its use among preschoolers is less common and rigorous studies among this age group are currently lacking.

The Survey of Wellbeing of Young Children (SWYC; Sheldrick, Henson, et al., 2012) is a comprehensive screening instrument for children under 5 years of age designed to be used easily in busy practice settings, particularly within primary care. There are 12 age-specific versions of the SWYC (2, 4, 6, 9, 12, 15, 18, 24, 30, 36, 48, and 60 months) that vary in question content. All versions of the SWYC include sections on developmental milestones (10 questions covering language, motor, social, and cognitive), family context (nine questions covering family risk factors such as parent depression and substance use), parent concerns (two questions), and child behavior. Regarding child behavior questions, parents of children up to 18 months children complete the Baby Pediatric Symptom Checklist (BPSC; 12 items) and parents of children 18 to 66 months complete the Preschool Pediatric Symptoms Checklist (PPSC; 18 items). The BPSC and the PPSC represent downward extensions of the Pediatric Symptoms Checklist, a behavioral screening tool that can be used with children 6 to 16 years of age. The SWYC also includes a seven-item autism-specific screener, the Parent’s Observations of Social Interactions (POSI), which parents of children 16 to 30 months of age complete. Sections of the SWYC that have been examined (BPSC, PPSC, POSI, milestones) demonstrate acceptable psychometric properties, including reliability (test–retest, internal consistency) and validity (Sheldrick, Henson, et al., 2012; Sheldrick et al., 2013; Sheldrick & Perrin, 2013; Smith, Sheldrick, & Perrin, 2013).

Comprehensive Dimensional Tools for Assessing Social-Emotional/Behavioral Problems

When the goal of the assessment is to gather detailed information about the specific areas in which a child is manifesting problems, longer, more detailed assessments are generally most appropriate. Comprehensive checklists can provide a detailed and differentiated understanding of the areas in which a child is reported by the parent to be manifesting a problem and those in which the child’s behavior falls in the typical range for his age group. These

types of tools ensure that parents are queried about a broad array of social-emotional problems including behaviors that parents may not be aware can be problems for young children. Comprehensive checklists usually gather information at the level of broad domains and subscales within each domain. The most common domains covered are the internalizing and externalizing domains first identified by Achenbach (Achenbach, 1966). Internalizing problems include fears, anxieties, and depressive and mood problems. Externalizing problems generally include aggressive, disruptive, and non-compliant behaviors, as well as problems of hyperactivity and inattention. Some tools, such as the ITSEA, also assess functioning in other domains, such as dysregulation and social competence (Table 1.7). Similarly, some tools assess functioning in a single domain; for example, the Multidimensional Assessment Profile of Disruptive Behavior (Wakschlag et al., 2014) focuses solely on ensuring a comprehensive assessment of disruptive behaviors, such as noncompliance, temper loss and low concern for others. In addition to providing domain-level scores that indicate a child's overall level of problems within a given broad area, these types of tools also provide subscale scores that indicate a child's functioning in more specific, narrowband areas of functioning. Clinically, having a profile that can pinpoint specific areas of difficulty can be more informative for guiding conversations with parents and treatment planning than broad domain scores. For example, on the ITSEA (Carter & Briggs-Gowan, 2006) and the CBCL/1.5-5 (Achenbach & Rescorla, 2000), the externalizing domain score is calculated by summing across a number of hyperactivity/impulsivity items as well as items describing aggressive and noncompliant behaviors. Given the frequent co-occurrence of disruptive and hyperactivity/impulsive behaviors clinically, a high score may reflect the presence of a wide array of problems throughout the domain. However, some children may display primarily hyperactive/impulsive problems without any elevation in aggression or noncompliance and others may behave in primarily disruptive ways with elevations in all areas. Thus, an elevated domain score provides an imprecise reflection of a child's problem behaviors. In contrast, subscale scores provide more specific information, enabling the clinical formulation to focus on the presenting problems.

Clinically, comprehensive checklists can be especially helpful when the goal is to profile a child's strengths and weaknesses to guide treatment planning. They also are best suited when the goal is to track treatment outcomes, because they provide a detailed assessment of functioning within the specific subareas that were identified as problems when the child began treatment and targeted by the

treatment plan. In contrast, using a tool that captures behavior only at the level of the broad domain runs the risk of underestimating improvement. This is because so many behaviors influence domain-level scores. Taking the externalizing domain as an example, if a child entered treatment with problems with hyperactivity/impulsivity and aggression and made significant gains in the area of aggression but not hyperactivity/impulsivity, the externalizing domain score could remain elevated above a clinical cutoff even though the child's aggressive behavior had dropped to normative levels.

Scientifically, increasing evidence suggests that narrowband subscales that capture discrete sets of behavior are more effective than scales that are comprised of heterogeneous sets of behavior for linking behaviors with neurobiological risk factors. For example, efforts to identify genetic contributions to psychopathology have been strengthened by the use of narrowband measures of phenotypic expression (Krueger, 1999; Krueger et al., 2002; Tackett et al., 2013; Todd et al., 2005). As an illustration in young children, one of our studies recently revealed that maternal reports of preschoolers' difficulty learning from punishment (or "Punishment Insensitivity" on the Multidimensional Assessment Profile of Disruptive Behavior) was associated with poorer performance on a computerized task that assessed learning from punishment (Briggs-Gowan et al., 2013). However, mothers' reports of other types of disruptive behavior were not associated with task performance. This narrowband approach also is consistent with the National Institute of Mental Health's Research Domain Criteria framework, which calls for research that employs dimensional measures to characterize aspects of functioning that can be more precisely linked with underlying biological processes, such as genetic vulnerability and neural pathways (Casey, Oliveri, & Insel, 2014; Insel et al., 2010). For these reasons, when the goals of an assessment are to dig more deeply in an effort to drive scientific discovery, comprehensive tools that provide both broad and narrow measures of functioning are often most desirable and a close inspection of the items that contribute to subscales and domains is essential to ensure that the tool captures the constructs of interest.

Selected Dimensional Checklists

Several norm-referenced tools for assessing social-emotional and behavioral problems common in this developmental period are summarized in Table 1.6 and 1.7 and described briefly in this section. These measures are fairly comparable in their general format (as shown

TABLE 1.6 Comprehensive Checklist/Questionnaire Measures That Address Social-Emotional and Behavioral Problems Broadly

| | ITSEA | CBCL | BASC-2 | PKBS-2 | Conners-EC Behavior Scales ¹ |
|-----------------------|---|--|---|--|--|
| | Infant-Toddler Social and Emotional Assessment (Carter & Briggs-Gowan, 2006; Carter et al., 2003) | Child Behavior Checklist ages 1.5-5 (Achenbach & Rescorla, 2000) | Behavior Assessment System for Children, Second Edition (Preschool Forms) (Reynolds & Kamphaus, 2004) | Preschool and Kindergarten Behavior Scales: Second Edition (Merrell, 2002) | Conners Early Childhood (Conners, 2009) |
| Age (months) | 12 to 36 months | 18 to 60 months | 24 to 72 months | 36 to 83 months | 24 to 83 months |
| Norming age bands | 6-month age bands | One norm group spanning 42 months | Separate norms for 2-3-year-olds and 4-5-year-olds | One norm group spanning 48 months | 6 month age bands |
| Gender-based norms | Y | Y | Y | N | Y |
| Description | Checklist | Checklist | Checklist | Checklist | Checklist |
| Respondent | Parent/teacher | Parent/teacher | Parent/teacher | Parent/teacher | Parent/teacher |
| Reading level | 4th to 6th | 5th | 4th | Not stated, but low | 5th |
| Number of items | 169 | 99 in parent | 134 in parent 100 in teacher | 76 in parent and teacher | 110 in parent and 112 in teacher |
| Administration time | 30 minutes | 15 to 20 minutes | 10 to 20 minutes | 10 to 15 minutes | 25 minutes |
| Time frame | Past month | Past two months | Last several months | Past 3 months | Past month |
| Response options | 3-point rating scale: Not true or rarely, somewhat true or sometimes, very true or often | 3-point rating scale: Not true, somewhat or sometimes true, very true or often true | 4-point rating scale: Never occurs, sometimes occurs, often occurs, almost always occurs | 4-point scale: Never, rarely, sometimes, often | 4 point rating scale: Not true at all (never, seldom), Just a little true (Occasionally), Pretty much true (Often, Quite a bit), Very much true (Very often, Very frequently) |
| Standardized scores | Y | Y | Y | Y | Y |
| Clinical cut points | Y | Y | Y | Y | Y |
| Available in Spanish? | Yes | Yes | Yes | Yes | Yes |

¹The Conners-EC also includes Developmental Milestones Scales (an additional 75 parent and 70 teacher/child-care items) for assessing Adaptive Skills, Communication, Motor Skills, Play, and Pre-Academic/Cognitive development.

in Table 1.6). They are written at a fourth- to sixth-grade reading level, are composed of 100–200 items, and require 15–30 minutes to complete. As normed instruments, these tools provide a means of comparing a given child with population-based means to determine whether the child's difficulties fall in a range that is considered clinically concerning. Despite these commonalities, these measures vary with respect to key factors including (1) whether they were explicitly developed for young children or are instead downward extensions of tools originally designed for older children; (2) the number and content of the domains they cover; and (3) the extent to which they provide insight into specific types of problems.

These tools have very detailed manuals with extensive information about the psychometric properties of each measure. Therefore, we do not attempt to summarize them

here. Overall, each of these tools *generally* has acceptable reliability and validity; however, because these are very comprehensive measures with many possible scores, it is possible that an individual scale or even a domain *may not have inadequate* reliability or validity in a particular population or age or sex group. Therefore, a careful review of a tool's reliability and validity within the population to be assessed is critical.

Comprehensive, Multidomain Checklists

The BASC-2 Preschool (Reynolds & Kamphaus, 2004) measures behavior problems and adaptive skills for children ages 2 to 5 years of age. The BASC-2 Preschool version is a parent or teacher report measure that is composed of 100–134 items depending on the informant. The reliability and validity of this measure have been

TABLE 1.7 Breadth and Depth of Coverage by Comprehensive Checklist Measures

| | ITSEA | CBCL | BASC-2 | PKBS-2 | Conners-EC |
|---|---|---|---|---|--|
| BREADTH OF COVERAGE: | | | | | |
| Cover a broad array of problems? | Y | Y | Y | Y | Y |
| Are there scores for multiple problem domains? | Y Internalizing, Externalizing, and Dysregulation Domains | Y Internalizing and Externalizing Domains | Y Internalizing and Externalizing Problems and Behavioral Symptoms Index | Y Internalizing and Externalizing Domains | N |
| Are there domain scores for social competencies and skills? | Y Competence Domain | N | N Adaptive Skills includes both social and nonsocial content | Y Social Skills Domain | Limited Social Functioning/ Atypical |
| Are there scores for autism spectrum problems? | Y Atypical and Social Relatedness Indices, Empathy, Prosocial Peer, Imitation/Play | Y DSM-Pervasive Developmental Problems | N Some relevant behaviors in Atypicality scale | N | N Some relevant behaviors in Social Functioning/ Atypical Behaviors |
| Are there scores for atypical or maladaptive behaviors? | Y Social Relatedness, Atypical, Maladaptive Indices | Y DSM-Pervasive Developmental Problems | Y Atypicality | N | N Social Functioning/ Atypical Behaviors |
| Are there scores for sensory or regulatory problems? | Y Sleep, Eating, Sensory Sensitivities, Negative Emotionality | Y Sleep Problems, Emotionally Reactive | N | N | Y Sleep Problems |
| DEPTH OF COVERAGE WITHIN SPECIFIC AREAS: | | | | | |
| Scale scores specifically about anxiety? | Y General Anxiety, Separation Distress, and Inhibition to Novelty Subscales | Y DSM-Oriented Anxiety Problems | Y Anxiety Scale | Y Anxiety/Somatic Problems | Y Anxiety |
| Scale scores specifically about depression or withdrawal? | Y Depression-Withdrawal | Y Withdrawn or DSM-Oriented Affective Problems | Y Depression, Withdrawal | Y Social Withdrawal | Y Mood and Affect |
| Combined attention and hyperactivity and impulsivity scores | N | Y DSM-Oriented Attention Deficit/Hyperactivity | N | Y Attention Problems/Overactive (note that this scale includes disruptive behaviors) | Y Inattention/ Hyperactivity |
| Scale score specifically about attention? | Y Attention Skills | Y However, Attention Problems includes hyperactivity "Can't sit still" | Y Attention Problems | N | N |

(continued)

TABLE 1.7 (Continued)

| | ITSEA | CBCL | BASC-2 | PKBS-2 | Conners-EC |
|--|--|--|----------------------------------|--|---|
| DEPTH OF COVERAGE WITHIN SPECIFIC AREAS: (continued) | | | | | |
| Scale score specifically about hyperactivity and impulsivity? | Y Activity/Impulsivity | N | Y Hyperactivity | N | N |
| Scale score specifically about disruptive behaviors | Y Aggression/Defiance, Peer Aggression | Y Aggressive Behavior or DSM-Oppositional Defiant | Y Aggression | Y Self-centered/ Explosive, Antisocial/ Aggressive | Y Defiant/Aggressive |
| Scale scores specifically about social skills and competencies | Y Empathy, Compliance, Prosocial Peer, Mastery Motivation, Imitation/Play | N | Y Social Skills, Adaptability | Y Social Cooperation, Social Interaction, Social Independence | Y Play subscale |
| Are there any questions to gauge parental worry, concern, or impairment? | Y Two questions gauge worry. They are not used in scoring | Y Open-ended question about presenting complaint | N | N | Y Parent rates how much the child's problems affect learning, peer relationships, and home life. |

evaluated. Norms are provided for 2- to 3-year-olds and 4- to 5-year-olds, with additional breakdowns for boys and girls, as shown in Table 1.7. The BASC-2 Preschool version provides four broad scores: Internalizing Problems, Externalizing Problems, Behavior Symptoms Index, and Adaptive Skills. It is important to understand the subscales that go into scoring each of these domains because some subscales are used in multiple domain scores. More specifically, the Externalizing score is based on the Aggression and Hyperactive scales. The Internalizing Problems score is based on Anxiety, Depression, and Somatic problems. The Behavior Symptoms Index is comprised of Hyperactivity, Aggression, Depression, Withdrawal, Attention Problems, and Atypicality. Thus, the Behavior Symptoms Index includes all Externalizing subscales, only one of three Internalizing scales, and three scales not included on either the Internalizing or Externalizing Problem scores. Therefore, it is important to be very careful to interpret patterns cautiously. For example, if the Externalizing and Behavior Symptoms scores are both elevated, it will be relevant to inspect the scales within each to determine whether both of these elevations are driven primarily by Hyperactivity and Aggression problems or whether other scales such as Attention, Withdrawal, Anxiety, or

Atypicality are also elevated and contributing to the high Behavior Problem Index score. Caution will need to be taken when using the BASC-2 in research applications to ensure that this overlap in scale content does not lead to misinterpretation of findings.

The Child Behavior Checklist for 1.5-5 (CBCL/1.5-5) (Achenbach & Rescorla, 2000) is a checklist measure that is a downward extension of the CBCL, originally developed for older children. The reliability and validity of the younger child version has been evaluated in children from 18 months through 5 years. The CBCL/1.5-5 addresses social-emotional/behavioral problems in three domains (Internalizing, Externalizing, and Total Problems), with more specific subscales within the Internalizing and Externalizing domains. The CBCL can also be scored to obtain *DSM*-oriented scales, which are designed to parallel symptoms that fall within specific diagnostic areas. It is worth comment that items that go into scoring the *DSM*-oriented scales also go into scoring the Internalizing and Externalizing Domain scores and subscale scores. Therefore, to avoid double counting individual problem behaviors, it is best to use one or the other set of scores but not both. Social-emotional competencies are not rated in

the CBCL/1.5-5. This measure also has been used internationally, including in Denmark (Kristensen, Henriksen, & Bilenberg, 2010), China (Liu, Cheng, & Leung, 2011), Taiwan (Wu et al., 2012), and Columbia (Flink et al., 2013).

The Conners Early Childhood (Conners-EC) (Conners, 2009) is a checklist measure that assesses both social-emotional and behavioral problems and developmental milestones in young children ages 2 to 6 years. Earlier forms of the Conners were developed for older children and focused on ADHD and other types of behavior problems (Conners, Sitarenios, Parker, & Epstein, 1998). The current version is extended to include anxiety, mood/affective problems, and atypical behaviors. The Conners-EC Behavioral form includes 115 parent-report and 116 teacher report items. It was normed within six-month age bands for boys and girls separately, which should enhance the developmental appropriateness and accuracy of its norms. This measure provides scores in six areas as follow: (1) Inattention/Hyperactivity; (2) Defiant/Aggressive; (3) Social Functioning/Atypical; (4) Anxiety; (5) Mood/Affective Problems; and (6) Physical Symptoms. The Conners-EC does not, however, offer broad domain scores for Internalizing or Externalizing Problems or for Social Skills/Competencies. However, the separate Developmental Milestones scales portion of the Conners-EC (eighty parent items, seventy-four teacher items) includes assessment of Play, as well as Adaptive behavior, Communication, and Motor Skills.

The Infant-Toddler Social and Emotional Assessment (ITSEA) (Briggs-Gowan & Carter, 1998; Carter & Briggs-Gowan, 2006; Carter et al., 2003) is a 169-item questionnaire about social-emotional and behavioral problems and competencies in 1- to 3-year-old children. From its inception, the ITSEA was developed for use with very young children. Items were developed based on literature review, clinical and developmental theory, and input from a large panel of experts from the fields of infant mental health, developmental psychology, and developmental psychopathology. It was developed as a comprehensive tool that would be appropriate for profiling children's strengths and weaknesses in the social-emotional domain. There are four broad domains, Internalizing, Externalizing, Dysregulation, and Competence, with multiple subscales within each domain. In addition, the ITSEA includes three indices, Social Relatedness, Atypical Behaviors, and Maladaptive. These indices consist of low base rate, but clinically significant behaviors. Reliability and validity have been established in four independent samples and in a national norming sample in the United States (Briggs-Gowan & Carter, 1998, 2008a; Carter &

Briggs-Gowan, 2006; Carter et al., 2003; Carter, Little, Briggs-Gowan, & Kogan, 1999). In international samples, a number of researchers have established the ITSEA's reliability and validity in other languages and cultures, including China (Jianduan et al., 2009), the Netherlands (Visser et al., 2010), and France (Bracha et al., 2004). Additional studies have shown that the ITSEA is sensitive to treatment change (Barlow et al., 2013; Lowell, Carter, Godoy, Paulicin, & Briggs-Gowan, 2011; Spittle et al., 2010).

The Preschool and Kindergarten Behavior Scales (PKBS-2) (Merrell, 2002) is a checklist measure of problem behaviors and social skills in 3- to 6-year-olds. This 99-item measure includes identical items for different informants. The exact same form is used for parents and teachers, but different norms are applied depending on the informant. The PKBS provides a somewhat less detailed set of scores about a child relative to other measures in this section. It offers Internalizing, Externalizing, and Social Skills scores. Standard scoring provides broad scores for the Internalizing, Externalizing, and Social Skills domains, as well as specific scores for three social skills subscales, Social Cooperation, Social Interaction, and Social Independence. Although they are not described in the scoring forms for the PKBS, the manual describes scoring for several subscales within the internalizing and externalizing areas, including Anxiety/Somatic Problems, Social Withdrawal, Attention Problems/Overactive, Self-Centered/Explosive, and Antisocial/Aggressive. A full understanding of the content of these scales is important however, as some scales include items that are not clearly described by the name of the scale. For example, the Attention Problems/Overactive scale includes some items that appear to be more disruptive in nature. Given the high levels of comorbidity within specific externalizing behaviors, it is likely that these items loaded best on this scale. However, interpretation should be informed by the content of the scales.

Dimensional Checklists That Focus on Specific Areas

In addition to previously described the comprehensive measures, a number of tools focus more narrowly on specific aspects of social-emotional/behavioral problems and competencies. The Preschool Anxiety Scale (Edwards et al., 2010; Spence et al., 2001) assesses anxieties in the areas of social anxiety, generalized anxiety, separation, and specific fears. The PAS has demonstrated fair to good internal consistency and factor structure (Broeren & Muris, 2008; Edwards et al., 2010; Spence et al., 2001). However, some scales have had poor internal consistency in some studies. For example, Separation Anxiety had

poor internal consistency in Chinese and Dutch samples (Broeren & Muris, 2008; Wang, 2014). As discussed earlier, this could reflect problems with the translation, cultural variation in how items are understood, or differences in our separation anxiety manifests in those cultures. The PAS has not to our knowledge been formally normed. Another tool, the Multidimensional Profile of Disruptive Behavior (MAP-DB) (Wakschlag et al., 2014), is a new 78-item parent report measure of disruptive behavior in the areas of Aggression, Temper Loss, Low Concern for Others (disregard, punishment insensitivity), and Non-compliance. The MAP-DB has demonstrated acceptable internal consistency (Cronbach's alphas of .92 to .97) and test-retest reliability (ICCs of .76 to .85) (Nichols et al., in press; Wakschlag et al., 2014). Evidence of validity has begun to emerge (Briggs-Gowan et al., 2013; Nichols et al., in press; Wakschlag, Choi, et al., 2012). A final example is the Social Skills Improvement System (SSIS) (Gresham & Elliott, 2008), a checklist measure for children ages 3 to 18 that emphasizes the assessment of social skills but also includes some scales for assessing problems in other areas (e.g., externalizing).

Variation in Emphasis of the Domains That Are Assessed

A review of the domains and subscales assessed by these different measures, summarized in Table 1.7, reveals variation in the areas assessed. There are three primary points to note.

First, most but not all of these measures provide domain-level scores for both internalizing and externalizing problems (Table 1.7). The Conners-EC does not provide summary scores at the broad domain level. While this may be appropriate in clinical settings, it may be a disadvantage for some research applications where it is sometimes advantageous to examine patterns at the broad domain level instead of (or in addition to) at the level of multiple subscale scores.

Second, only some of these measures assess social skills/competencies and there is considerable variability within these in the types of competencies that are assessed. The BASC-2, ITSEA, and PKBS all assess social competencies and include subscales that assess specific aspects of social-emotional competence. For example, the PKBS assesses skills in the areas of social cooperation, social interaction, and social independence, the BASC captures social skills and adaptability, and the ITSEA provides scores for empathy, compliance, prosocial peer play, mastery motivation, and imitation/play. In contrast, the CBCL/1.5-5 does not assess any social competencies.

Third, there is considerable variability across these measures in their coverage of behaviors that may be associated with autism spectrum disorder. For example, the ITSEA includes indices designed to characterize these types of problems and has also demonstrated an empirical profile that is associated with autism (Carter & Briggs-Gowan, 2006). This profile is characterized by high scores on the Atypical Index and Depression/Withdrawal and low scores on the Social Relatedness Index and the Imitation/Play and Empathy. Similarly, the CBCL/1.5-5 has identified a Pervasive Developmental Delay subscale for assessing problems in this area. Other measures, such as the Conners-EC, have identified profiles that are associated with pervasive developmental delays. As a profile of specific weaknesses may be expected with children who are on the autism spectrum, a careful review of the manual is warranted to ensure appropriate interpretation of the profile of findings, is important when ASD is a consideration in assessment.

Finally, these measures differ in how deeply and specifically particular types of problems are assessed. ADHD problems provide a good example. Two measures have separate subscales for attention problems and hyperactivity/impulsivity problems, whereas the remaining measures combine these behaviors into a single score. For anxiety, most of the measures have a single subscale for assessing anxiety or combine anxiety and somatic problems; the latter approach can be problematic if working with medically ill children. The ITSEA is unique in distinguishing different types of anxiety, with scales for separation distress and general anxiety.

A final point to consider is variation in how child age and sex were addressed in the norming process. For example, the Conners-EC and ITSEA derived norms for six-month age bands for boys and girls separately and thus provide very narrow comparison groups against which children are compared. In contrast, the CBCL and PKBS employ the same norms across the entire age group for which the tool is assessed and did not derive norms separately for boys and girls. The BASC-2 used an intermediate approach, with age bands that span 24-month periods for boys and girls separately.

Diagnostic Approaches

Diagnostic interviews are designed to gather the detailed information that is needed to establish whether all of the criteria for making a particular diagnosis are met. The criteria required for a diagnosis varies across disorders, but often includes a specific number of symptoms, a period of duration, age of onset, and impairment. Although mostly

used in research settings, structured and semistructured interviews with parents or caregivers offer a more in-depth examination of child problems than can be obtained from questionnaires and can be particularly useful for helping to determine the child's diagnostic status (Egger & Angold, 2004). For example, in the PAPA interview, described in detail in the following section, interviewers can obtain specific examples of child behavior and inquire about the onset, offset, frequency, duration, intensity, quality, and contexts in which behaviors occur through follow-up questions and probes. Thus, interviews can provide a means of obtaining a more comprehensive understanding of the child's behavior and gathering information necessary to assign diagnoses.

The decision to use diagnostic interviews in the assessment process brings added challenges, including the staff and family time for the interview itself and staff time for review and scoring to establish diagnostic status. Interviews also typically require a formal training in administration and ongoing monitoring of the fidelity of administration and any ratings that are made in the scoring process to minimize drift over time. Such procedures are crucial to ensure that children are being consistently and systematically assessed in the same way. In practice, however, these standards can be difficult to uphold, especially when funding is limited or staff change over time, necessitating a system for training new staff. These issues are pronounced in early childhood because all of the interviews that are available are semistructured. Semistructured interviews provide interviewers with a clear framework for the interview, but interviewers must use their judgment about what constitutes a symptom, when to probe further to establish symptom presence, and what phrasing to use to describe a symptom (Angold et al., 2012). While the interviews vary in the background training required, from lay interviewer to clinician, the fact that interviewer judgment is employed means that training and monitoring are always essential.

The reliability of an interview provides information about how consistently the interview elicits the same information. Test–retest reliability indicates how well the interview provides the same results when administered twice at a short enough interval (typically 1–4 weeks) that true change in the child is not likely to have occurred. As it is administered by different interviewers at the two time points, test–retest reliability can be affected by differences in how deeply each interviewer probes or differences in how they perceive the information obtained even if they probe exactly the same way. Test–retest reliability can be influenced by true change in the child over time and development. Changes in how the parent perceives the

child's behavior, fluctuations in the parent's attention level or mood can also affect reliability. The experience of being interviewed may influence a parent's understanding of the child's behavior. For example, they may come to realize that some behaviors that they thought were normal may be symptomatic or they may watch their child's behavior and realize it occurs more or less frequently than they thought it had at the first interview. Parents often also learn that if they deny the presence of symptoms the interview will go faster, resulting in lower scores at the second interview. Indeed, Egger and her colleagues note that attenuation (reduction) of symptoms and disorders is a common problem in test–retest studies of diagnostic interviews (Egger et al., 2006). Interrater reliability is a second type of reliability that's important for interviews. It refers to how consistently different people rate the same interview. Interrater reliability is typically assessed by having raters watch videotapes of the same interview and code it independently. It also can be assessed by having one rater administer and code the interview and the second rater code from an audio or videotape. A constraint of inter-rater reliability it does not establish how well the interview was administered, only whether the information obtained yields comparable diagnostic and symptom data when rated independently. Together, test–retest and interrater provide complementary information about reliability. Internal consistency can also be assessed for symptom scores obtained from interviews, but this is often not reported.

Selected Diagnostic Interviews

In this section, we will review and discuss four interviews that have been used with young children (Table 1.8). First we will review two interviews that were explicitly developed for early childhood: the Preschool Age Psychiatric Assessment (PAPA) (Egger & Angold, 2004; Egger et al., 2006) and the Diagnostic Infant and Preschool Assessment (DIPA) (Scheeringa, 2004; Scheeringa & Haslett, 2010). Then we will review two interviews originally developed for older children—the Anxiety Disorders Interview Schedule (ADIS) (Silverman, Saavedra, & Pina, 2001) and the Schedule for Affective Disorders and Schizophrenia – Present and Lifetime (K-SADS-PL) (Kaufman et al., 1997; Kaufman, Birmaher, Brent, Rao, & Ryan, 1996)—which have been used in studies of preschool-age children.

PAPA

The PAPA (Egger et al., 1999) was developed for parents of 2- to 5-year-olds. It is designed to provide comprehensive and developmentally sensitive coverage of the

TABLE 1.8 Diagnostic Interviews Used with Young Children

| | PAPA (Egger et al., 2006) | DIPA (Scheeringa & Haslett, 2010) | K-SADS-PL Preschool (Kaufman, Birmaher, Brent, Rao, & et al., 1997) | ADIS (Silverman et al., 2001) |
|--------------------------------|--|--|--|--|
| Age (months) | 2 to 5 years | <7 years | 1 to 6 years | 3 to 5 years |
| Interviewer type | Lay or clinician | Lay or clinician | Clinician recommended | Clinician |
| Interview format | Semistructured | Structured and Semistructured | Semistructured | Semistructured |
| Training required | Yes | Yes | Yes | Yes |
| Respondent | Parent | Parent | Parent | Parent |
| Time frame covered | Past 3 months (and lifetime as appropriate) | Past 4 weeks | Past 3 months (and lifetime as appropriate) | Not specified |
| DSM version | <i>DSM-IV-TR/RDC?</i> | <i>DSM-IV</i> | <i>DSM-IV</i> | <i>DSM-IV</i> |
| Scoring | Electronic database | Hand scoring | Diagnostic consensus or by expert review | Hand scoring |
| Developmental modifications | Developed for young children | Developed for young children | Developed for older children. Studies vary in their developmental modification. | Developed for older children. Developmentally inappropriate questions are excluded. |
| Diagnostic areas covered | Common childhood disorders | Common childhood disorders | In preschoolers, reliability and validity primarily available for anxiety, ODD, CD, and ADHD | Anxiety disorders, ADHD, ODD, sleep terror disorder, selective mutism |
| Impairment | Impairment in multiple domains is evaluated in an incapacity module at the end of the interview | Impairment in multiple domains is evaluated at the end of each diagnostic section | Assessed in the process of assessing symptoms | Impairment in multiple domains is evaluated at the end of each diagnostic section |
| Test–retest reliability data | Reported (Egger et al., 2006) | Reported (Scheeringa & Haslett, 2010) | Reported for ODD/CD (Keenan et al., 2007) | Not reported for young children |
| Interrater reliability data | Not reported | Not reported | Reported for ODD/CD (Keenan et al., 2007) and for anxiety, ADHD, ODD, Mood (Birmaher et al., 2009) | Reported (Edwards et al., 2010; Rapee et al., 2005; Shamir-Essakow et al., 2005) |
| Validity data | Not explicitly studied, but various studies provide support for validity | Construct validity (Scheeringa & Haslett, 2010) | Reported for ODD/CD (Keenan et al., 2007) and for anxiety, ADHD, ODD, mood (Birmaher et al., 2009) | Not explicitly studied |

symptom, frequency, duration, and onset criteria necessary to generate diagnoses for the *DSM-IV*, ICD-10, and DC: 0-3 diagnostic classification systems. In developing the PAPA, the authors conducted thorough reviews of the infant mental health literature and solicited detailed input from many infant mental health experts (Egger & Angold, 2001). The PAPA assesses a broad set of internalizing disorders, externalizing disorders, and regulatory disorders (addressing, e.g., sleep, feeding, and emotion regulation) as well as PTSD and reactive attachment disorder. At the end of the interview, an “incapacity” section is used to determine the level of impairment that is associated

with symptoms of each diagnostic area. Impairment can be endorsed in several areas, including in parent–child and sibling relationships, in school or early care and education settings, and with peers. The PAPA may be administered by clinicians and laypersons, provided they have received formal PAPA training (Egger & Angold, 2004). Notably, the PAPA includes many adaptations that enhance its developmental appropriateness for young children. For example, it includes an extensive set of questions about temper tantrums, including probes to establish the types of behaviors that are manifested during tantrums (e.g., stamping feet, throwing things, holding breath), the

contexts in which they occur (home, school, elsewhere), how long they last, and whether the parent can identify triggers for tantrums (e.g., frustration, fatigue, transitions) or reports that they are out of the blue. A unique feature of the PAPA is that in determining whether symptom criteria are met, the diagnostic scoring algorithm incorporates information about the frequency of individual symptoms in a normative population. Thus, the algorithms incorporate empirical data to help distinguish normative from non-normative behaviors. Given that parents and professionals often have difficulty knowing when a particular behavior is problematic, this is a potentially valuable feature of the PAPA.

DIPA

Like the PAPA, the Diagnostic Infant and Preschool Assessment (DIPA) (Scheeringa, 2004; Scheeringa & Haslett, 2010) also covers a full range of psychiatric disorders that are common in childhood. It was designed for parents of children from infancy through 6 years of age. It is intended to provide a relatively quick and practical method for assessing psychiatric disorders. The DIPA is both structured and semistructured. It is structured in the sense that the initial stem questions about a given symptom are read verbatim, but semistructured in that the interviewer must probe for examples to establish whether symptom criteria are met. The interview is laid out in such a way that the interviewer can determine whether diagnostic criteria are met without requiring time-consuming data entry or computerized scoring. The DIPA also has incorporated a number of helpful developmental adaptations. For example, to aid the parent and interviewer in establishing whether a child's sleep patterns are symptomatic, the sleep disorders section incorporates empirical information for children of different ages (12 to 24 months versus 24 months and older) concerning how long children in this age group typically take to fall asleep and how often they typically waken during the night.

Two interviews that were originally developed for older children, the ADIS (Silverman et al., 2001) and the K-SADS-PL (Kaufman et al., 1996, 1997) have been used in studies of preschool-age children (Birmaher et al., 2009; Keenan et al., 2007; Kennedy, Rapee, & Edwards, 2009; Rapee, Kennedy, Ingram, Edwards, & Sweeney, 2005; Shamir-Essakow, Ungerer, & Rapee, 2005). Because they were developed for older children, in their original forms, these interviews include some *DSM* symptom criteria and descriptors that are developmentally inappropriate for young children, such as sexually inappropriate touching, vandalism, and stealing (Wakschlag et al., 2010). Most

studies that have used the ADIS and K-SADS with young children have simply excluded these developmentally inappropriate symptoms from the interview entirely (Birmaher et al., 2009; Shamir-Essakow et al., 2005). However, this approach may miss the opportunity to assess these symptoms in a developmentally meaningful and valid way. For example, Wakschlag and Keenan have shown that many disruptive behavior symptom criteria of the K-SADS can be operationalized in a developmentally appropriate way for preschool-age children. When they used the K-SADS disruptive behavior modules with preschool children, they excluded inappropriate items (e.g., forced sexual contact) and modified definitional examples of symptom criteria where developmentally relevant (Keenan et al., 2007). For example, an exemplar for "often bullies" could include threatening aggression to get a toy (Gray & Wakschlag, in press). As interviews vary in how they have handled these issues of developmental relevance, it is important to review the context of an interview carefully when deciding whether it suits your goals and when interpreting results.

Psychometric Properties of Diagnostic Interviews

A review of the literature reveals unevenness across interviews in the types of reliability that have been emphasized and reported. For example, test–retest reliability has been systematically assessed for the PAPA and DIPA, has been reported for only Disruptive Behavior Disorders in the K-SADS, and has not been reported for the ADIS. In contrast, interrater reliability has been reported in multiple studies that have used the ADIS with preschoolers.

Test–Retest Reliability

Our literature review identified studies that have reported test–retest reliability statistics in young children for the PAPA, DIPA, and the K-SADS Disruptive Behavior Disorders (K-DBDs) but not for the ADIS. As summarized in Table 1.9, this review suggests that test–retest reliability is generally better for symptom scores than for presence versus absence of disorder. The poorer reliability for disorders relative to symptoms may be explained by the fact that diagnostic status is less forgiving of modest change. For children whose difficulties fall right near the threshold for disorder, a minor change in number of symptoms can dramatically change diagnostic classification (from present to absent or vice versa) but have minimal impact on symptom scores. A detailed review of each interview is provided in the following sections.

TABLE 1.9 Test–Retest Reliability of Diagnostic Interviews Used with Young Children

| | PAPA (Egger et al., 2006) | | DIPA (Scheeringa & Haslett, 2010) | | K-SADS (Keenan et al., 2007) | |
|----------------------|-----------------------------------|--------------------------|-----------------------------------|--------------------------|------------------------------|------------------|
| | Disorders (Kappa) | Symptom | Disorders (Kappa) | Symptom | Disorders (Kappa) | Symptom |
| | Normative Sample/ Screen Highs | Scores (ICCS) | With/ Without Impairment | Scores (ICCS) | | Scores (ICCS) |
| INTERNALIZING | Poor to Excellent | Good | Fair to Good | Fair to Excellent | | |
| SAD | .60/.52 | .63/.68 | .50/.44 | .78 | - | - |
| GAD | .39/.34 | .61/.61 | Too rare | .41 | - | - |
| Specific phobias | .36/.49 | .57/.51 | - | - | - | - |
| Social phobia | .54/.63 | .73/.75 | - | - | - | - |
| PTSD | .73/.79 | .56/.71 | .56/.67 | .87 | - | - |
| MDD | .72/.61 | .71/.69 | Too rare | .40 | - | - |
| EXTERNALIZING | Fair to Good | Good to Excellent | Fair to Good | Good to Excellent | Good to Excellent | Excellent |
| ADHD | .74/.67 | .8 | - | - | - | - |
| ADHD–inattentive | - | - | .58/.78 | .73 | - | - |
| ADHD–hyperactive | - | - | .42/.47 | .65 | - | - |
| ODD | .57/.44 | .67 | .38/.47 | .78 | .81 | .98 |
| CD | .60/.54 | .66 | Not studied | Not studied | .73 | .81 |

Note: SAD=Separation Anxiety Disorder; GAD=Generalized Anxiety Disorder; PTSD=Post-Traumatic Stress Disorder; MDD=Major Depressive Disorder; ADHD=Attention-Deficit/Hyperactivity Disorder; ODD=Oppositional Defiant Disorder; CD=Conduct Disorder

PAPA. The reliability of the PAPA was examined in a sample of 2- to 5-year-old children who were recruited from a normative, population-based sample (Egger et al., 2006). To ensure the presence of sufficient psychopathology to evaluate the PAPA's reliability, children with high scores on the CBCL/1.5-5 were oversampled. Reliability is reported both for these screen-high children, as an estimate of reliability in clinical populations, and in a subsample that was more similar to the original normative sample as an estimate of reliability in nonreferred populations. The PAPA generally demonstrated good test–retest reliability for disorders (Table 1.9). Of 18 kappas reported, nine fell in the “Good” range (.60 to .74) and six were “Fair.” Three of eight kappas for anxiety disorders were “Poor.” In discussing this low reliability, Egger suggested that this could reflect poor reliability of *DSM* anxiety disorders in young children (Egger et al., 2006). All symptom count variables were reported to have good to excellent reliability. As noted by Egger and her colleagues (2006), the test–retest reliability of the PAPA is comparable to what has been reported for diagnostic interviews of older children and adolescents.

DIPA. The test–retest reliability of the DIPA was evaluated in a sample of 50 outpatient children under the age of 7 (Scheeringa & Haslett, 2010). Review of the DIPA revealed generally fair to good reliability for separation anxiety, PTSD, ADHD, and Oppositional Defiant

Disorder (Table 1.9). However, reliability could not be adequately assessed for other disorders because some disorders occurred too rarely to estimate reliability; a rule of thumb is that reliability should be estimated only if there are at least five children with the disorder in the sample. Specific and social phobias were not assessed due to time constraints and concern about burden on families and interviewers in this clinical sample. At the symptom level, similar to the PAPA, reliability for assessing ODD and ADHD symptoms was good to excellent. Reliability was excellent for Separation Anxiety and Post-traumatic Stress symptoms but fair for symptoms of Major Depression and Generalized Anxiety.

K-SADS. The test–retest reliability of the K-DBDs was evaluated in a study of 223 clinically referred and nonreferred children, ages 3 to 5 years (Keenan et al., 2007). The focus of this study was specifically disruptive behavior disorders. Test–retest reliability statistics for ODD and conduct disorder (CD) and symptoms were good to excellent (Table 1.9). To our knowledge, test–retest reliability has not been reported for other disorders assessed by the K-SADS in preschoolers.

Interrater Reliability

Much less information is available about the interrater reliability of these diagnostic interviews. A number of studies that have used the ADIS with preschool samples have

reported its interrater reliability in assessing presence or absence of anxiety disorders (Edwards et al., 2010; Rapee et al., 2005; Shamir-Essakow et al., 2005). Across these three studies, interrater reliability was consistently good to excellent for separation anxiety ($K = .69$ to $.89$), specific phobias ($K = .69$ to $.83$), and social phobia ($K = .84$ to $.89$). However, interrater reliability of generalized anxiety disorder has been only fair in two studies ($K = .54$ to $.56$) and excellent in one ($K = .77$). For the K-SADS-DBDS, Keenan and her colleagues (2007) reported excellent interrater reliability for ODD and CD symptoms (both ICCs = $.98$) and disorders ($K = .96$ and $.98$, respectively). Finally, although specific details were not provided for individual disorders, Birmaher and colleagues reported that the K-SADS demonstrated excellent interrater reliability for all disorders (all kappas $>.80$) in a study of preschool-age children (Birmaher et al., 2009).

Validity

Validity of diagnostic interviews in young children has rarely been tested explicitly. As a consequence, their validity in properly classifying children as having or not having psychiatric disorders is essentially unknown. This is at least partially a reflection of an inherent challenge in defining an appropriate gold standard for evaluating the validity of psychiatric diagnoses. While clinician-assigned diagnoses might be considered for this purpose, they can have poor reliability (Fisher et al., 1993). Thus, poor validity could arise from unreliability of the clinician diagnoses rather than a limitation of the interview. One approach to validate these measures is to demonstrate expected associations with checklist measures. This approach validates that the interview captures elevations in symptoms—it cannot validate that the diagnosis itself is valid because checklists cannot determine the presence or absence of disorder explicitly because they do not evaluate diagnostic criteria. Another approach is to demonstrate concordance between the interview and overall clinical status, that is that children referred for treatment are more likely to meet criteria for disorder than healthy controls.

The DIPA is unique in that Scheeringa and colleagues explicitly evaluated the validity of the continuous symptom scores in relation to the CBCL 1.5-5 (Scheeringa & Haslett, 2010). Results of these analyses indicated good validity for the externalizing symptom scores and for separation anxiety, with correlations from $.44$ to $.59$, but generally low validity for internalizing problems, such as generalized anxiety and depression, with most correlations below $.25$. However, this might be as much a limitation of the CBCL

as it is of the DIPA. For the K-SADS-DBDS, ODD and CD diagnoses demonstrated good reliability relative to measures of impairment and referral status (Keenan et al., 2007). To our knowledge, the validity of the PAPA, ADIS, and other diagnostic areas of the K-SADS have not been studied explicitly.

Nonetheless, as researchers employ these interviews in their studies, we have opportunities to gain insights into validity. Various studies that have used these interviews suggest their value for characterizing preschool psychopathology. For example, the PAPA, ADIS, and K-SADS all have been used in multiple preschool studies for a variety of purposes, including as a means of validating another tool (Edwards et al., 2010; Wakschlag et al., 2008) and of documenting treatment outcomes (Comer et al., 2012; Rapee et al., 2005) and for characterizing psychopathology in young children (Briggs-Gowan et al., 2010; Bufferd, Dougherty, Carlson, & Klein, 2011; Shamir-Essakow et al., 2005; Sterba, Egger, & Angold, 2007). For example, one recent study demonstrated greater impairment in young children with anxiety or depressive disorders identified by the PAPA relative to healthy children without disorders (Otto, Andreas, von Klitzing, Fuchs, & Klein, 2014). Another demonstrated the predictive validity of irritability, assessed with the PAPA, for predicting later depression, oppositional defiant disorder, and functional impairment (Dougherty et al., 2013). Longitudinal prediction from PAPA diagnosis to school-age peer impairment also has been reported (Danzig et al., 2013). Yet another demonstrated significant functional impairment in children who meet criteria for disorders on the PAPA (Bufferd et al., 2011). Treatment studies that have used the ADIS to assess anxiety disorders with preschool-age children have been similarly fruitful, with evidence that the ADIS is sensitive to treatment effects on anxiety disorders (Kennedy et al., 2009; Rapee et al., 2005). While these studies do not explicitly demonstrate the validity of these interviews, they do suggest their value for studying emergent symptoms and disorders in this developmental period.

It is worth commenting that a consistent emerging pattern is that generalized anxiety disorder is less reliably assessed, with lower test–retest and interrater reliability in most studies. This may reflect a need for further research to establish the validity of this type of anxiety in young children. This type of anxiety also may be difficult to assess in young children because their more limited ability to verbalize their internal states may mean that parents and other caregivers are less consistently aware of worrying and apprehension in this age group.

Observational Assessment

Observation is routinely used to assess the quality of individual, dyadic, triadic, and family functioning in young child assessments. These methods include two components: the procedures that are used to create the context for the observational episodes, and the coding system designed to quantify the elements of interest. Interactional episodes vary along a spectrum from naturalistic observation to structured interactions. Coding systems vary from micro-analytic coding of discrete behaviors across very brief time segments of interaction to global coding that requires appraisal of multiple components of the interaction across the entire duration of the observation.

A number of methods for assessing child temperament originated from developmental psychology laboratories. These include behavioral inhibition paradigms in which children are observed in situations in which they are exposed to novel objects and social situations (Kagan, Reznick, Clarke, Snidman, & Garcia-Coll, 1984; Reznick et al., 1986). The more extensive Laboratory Temperament Assessment Battery (Gagne, Van Hulle, Aksan, Essex, & Goldsmith, 2011) is a series of 3–5-minute naturalistic episodes that include the introduction of specific presses designed to capture variation in emotional expression, activity level, approach and avoidance, and other regulatory behaviors. Other measures focus on dyadic or triadic interactions. For example, the Parent-Child Early Relational Assessment (PCERA) is designed to assess individual child and parent affective and behavioral patterns and dyadic behaviors in both unstructured free play episodes and in more structured interactions (e.g., reading a book, doing a mildly difficult task) Clark, Tlucek, & Gallagher, 2004. Observations of triadic interactions, in which aspects of caregivers, cooperation, competition, and conflict are rated in relation to a particular child, have become more common in recent years (McHale & Coates, 2014; McHale & Fivaz-Depeursinge, 1999). The Disruptive Behavior Diagnostic System (DB-DOS) (Wakschlag et al., 2007, 2008) is an example of an observational paradigm and coding system that were developed to characterize children's functioning on a clinical continuum from normal to atypical. The interactions include parent-child and examiner-child interactions. Building on methods used in the area of ASD (Lord et al., 1989, 2000), each interactional context is designed to press for the full range of behaviors relevant to preschool disruptive behavior. For example, episodes include a don't touch shelf, a (presumably boring) sorting task, and a frustrating puzzle. As the presses build across the contexts, they can elicit valuable insight into how

a child responds when stressed or frustrated. For example, providing qualitative and quantitative information about the intensity, pervasiveness, and ease of escalation of negative affect as well as whether it is displayed with both the parent and the examiner. We have recently developed and reported the preliminary reliability and validity of an extension of this method to anxiety, the Anxiety Dimensional Observation System (Anx-DOS) (Briggs-Gowan, Mian, Carter, & Wakschlag, 2011; Mian et al., in press).

Observational methods such as these offer an important means of capturing independent information about a child's functioning. In practice, however, they are typically costly and burdensome to implement. To be valid, the observational paradigm must be administered in a standard fashion. Thus staff must be well trained, and there should be ongoing quality control checks to ensure that they continue to administer the protocol reliably over time. Similarly, observational coders must be trained to apply the coding system reliably, and their reliability must be maintained over time. The process of training a coder to the point of acceptable reliability can require many hours of coding and feedback with an expert coder. Once a coder is reliable, ongoing monitoring of reliability is essential to prevent drift over time. Consequently, standardized, coded observational approaches are primarily used in research settings. For coding of discrete behaviors, new technological advances are developing in which computers and computer software can assume some of the burden of coding; for example, improvements in face recognition and machine learning enabling coding of discrete emotions such as pain (Lucey et al., 2011).

Finally, along with observation, several researchers have developed methods for directly eliciting information about symptomatology from children. A variety of measures have been developed focused on the free responses of children to story stems, puppet interactions (Measelle, Ablow, Cowan, & Cowan, 1998; Warren, Emde, & Sroufe, 2000; Warren, Oppenheim, & Emde, 1996), and simple picture-based questionnaire responses (Ialongo, Edelsohn, Werthamer-Larsson, Crockett, & Kellam, 1993, 1995; Martini, Strayhorn, & Puig-Antich, 1990). However, in relation to other diagnoses, the questions of whether and how to incorporate observational measures, or such material elicited directly from the child, have not been settled.

Assessing Impairment

As in older children and adults, social-emotional and behavioral problems in young children can be associated with impairment in functioning (Briggs-Gowan &

Carter, 2008a; Briggs-Gowan et al., 2001, 2006; Egger & Angold, 2006; Keenan et al., 2007; Lavigne et al., 1996; Luby, Belden, Pausch, et al., 2009). Symptoms of psychopathology can interfere with a child's participation in age-appropriate activities, developmental progress or relationships with others. When evaluating impairment in a young child, it is important to consider how the child's developmental capacities in domains such as cognition, motor, or language affect functional impairment. Delays in other areas, such as language development, may impede progress in other domains. Consider for example, the challenges faced by a child with mixed language delays and gross motor delays when attempting to keep up with peers. It is important that challenges in functioning associated with the mixed language and gross motor delays not be attributed to the symptoms of psychopathology that the child may also be evidencing. Difficulties being understood by and understanding peers may impede opportunities to learn age-appropriate prosocial skills acquired through give and take with peers. Difficulties running and climbing may interfere with his efforts to join gross motor peer play, contribute to isolation, and interfere with the development of peer relationships. Thus, assessing impairment associated with psychopathology requires consideration of developmental trajectories in multiple domains and their impact on children's performance of daily routines and activities.

In older children and adults, impairment is typically assessed by evaluating the individual's interpersonal, academic, occupational, and adaptive functioning (Association, 2000; Organization, 2001). In contrast, working with young children necessitates shifting the emphasis from the child only to encompass both the child and the broader family system.

Very young children's functioning is dependent on parental scaffolding and support. Impairment may manifest in the extent to which parents make accommodations that ultimately impact their child's participation in age-appropriate activities or developmental progress. For example, although a parent may not be able to offer examples of *child* impairment, *the parent* may have quit his or her job and restructured the household to accommodate their child's special needs or social-emotional problems. The parent may also avoid situations that would trigger their child's problem behavior. For example, a parent with a very active child may refrain from taking her even to family-friendly restaurants or a parent with a socially anxious child may avoid even child-friendly activities in the community. Such minimization of the demands placed on the child may help to obscure the child's impairment (Clark et al., 2004) while at the same time limiting the

child's ability to engage in and learn from age-appropriate activities and likely further compounding the problem. When conceptualized within a broader family framework, young child impairment may manifest as parental distress or low efficacy about the parent-child relationship or evidence that the child's behavior interferes with the parent's ability to maintain family routines (e.g., have family meals together, visit relatives), household activities (e.g., making phone calls), or employment (e.g., restrictions on work due to the challenges of obtaining appropriate care for the child) (Clark et al., 2004).

Thus, when evaluating functional impairment, it is often useful to consider how a child's social-emotional/behavioral problems may limit the child's participation in age-appropriate activities such as learning activities (e.g., listening to books, drawing, building); opportunities to acquire and practice social skills with peers and nonfamilial adults; and exposure to challenging situations key to practicing and acquiring self-regulation skills. In parallel with this, it may be very illuminating to query how the child's problems may strain the parent-child relationship, cause distress for child, parent, or other family members, interfere with family routines (meals, bedtime, getting ready for the day), impact siblings, limit parental employment, or be perceived as affecting parental well-being (e.g. lack of sleep, worry, fatigue) or straining family relationships (between parents or in the broader family context).

Many available assessments now include some gauge of impairment in child or family functioning, as noted in the aforementioned summary of assessments. In addition to these, a number of measures offer a means of gathering information about the impact of the child's behavior on family functioning, such as the Parenting Stress Index (Abidin, 1990) and the Family Life Impairment Scale (Briggs-Gowan, Horowitz, & Carter, 1998; Briggs-Gowan et al., 2006; Carter et al., 2010). Measures of adaptive behavior, such as various forms of the Vineland Adaptive Behavior Scales II (Sparrow, Cicchetti, & Balla, 2005) and the Adaptive Behavior Assessment System (ABAS-II) (Harrison & Oakland, 2003) are norm-referenced tools that provide comprehensive assessment of a child's adaptive functioning. Adaptive functioning that is significantly lower than cognitive functioning often reflects the child's inability to participate in age-expected activities or attain developmentally expected competencies. Thus, adaptive functioning may represent one important index of impairment.

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

There has been tremendous progress in developing new instruments that enhance assessment of young children's social-emotional and behavior problems and disorders. Specifically, an array of instruments now exist: parent-reported questionnaires are available for screening and more comprehensive assessments of social-emotional and behavior problem and competence dimensions; diagnostic interviews are available for parents of children 2 years and older; and observational assessments of temperament and specific symptoms (e.g., disruptive behaviors) have been shown to deepen understanding of children's development in relation to their peers. Yet, despite these efforts and successes, many challenges remain. First, integrating data across informants and methods continues to be a significant challenge in both clinical and research settings, particularly given the frequent occurrence of low correlations between informants and methods (i.e., questionnaires, interviews, and observations). Given that young children do not have the metacognitive capacities to report on their own symptoms until approximately 4 years of age (and then only when using developmentally appropriate interviews such as the Berkeley Puppet Interview (Measelle et al., 1998), greater variability in caregiver expectations for child behavior among very young children, meaningful variability in children's behavior in different contexts (Dirks et al., 2012), and lack of consistency in documenting impairment among young children, recommendations or guidelines for systematic and valid approaches for integrating data would be a major contribution to the field. Recommendations should be based on empirically supported methods for integrating of multimethod, multi-informant assessment data within the domain of social-emotional and behavior problems and disorders that often yield competing views of the child. Moreover, such methods should also address children's developmental functioning and contextual factors such as caregiver-child relationships, family functioning, culture, and broader contextual factors that may shape the behaviors and emotions that parents and other informants notice and the meanings that they attribute to these behaviors and emotions.

Second, with an increased push for screening children for social-emotional and behavioral problems early and often, it becomes imperative to know how repeated administration of the same instrument may influence psychometric properties over time. One could easily hypothesize attenuation effects due to repeated screening

as parents pay less attention to items, or one could hypothesize increased rates of reporting as parents become familiar with items and may begin to pay more careful attention to the child behaviors associated with screening items. Repeated screening also offers an opportunity for developing new methods for determining screening thresholds. Specifically, it is plausible that a child whose screening score is at or is one point above the single time-point threshold score on a screening instrument should carry more weight in relation to follow-up assessment or referral if similarly elevated scores have been obtained on two earlier occasions and less weight if this is the first occurrence of a score that just exceeds the cut score that is routinely employed for this age group. Underscoring the potential value of new methods for integrating longitudinal screening information, emerging research has shown that repeated assessments of irritability enhance prediction of impairment in preschoolers (Wakschlag et al., in press).

Third, with increased global diversity, there is a striking need to improve the availability of social-emotional and behavioral measures that have culturally valid content and offer representative normative data that are appropriate to the diversity that currently exists within many countries. This work will likely require international collaborative efforts using mixed methods approaches (i.e., integrated qualitative and quantitative studies) to aid instrument developers to serve the diverse children and families whose children may be at risk for or exhibiting elevated social-emotional and behavioral symptoms or disorders. Excellent translations of instruments are insufficient to meet this challenge. We need information about the psychometric properties of widely used measures in the many cultural groups in which we are currently employing translated tools. Indeed, there may be opportunities to capitalize on ongoing large-scale longitudinal epidemiologic studies in various countries to establish cross-national norms and validation. Moreover, given the diversity within individuals who share a common language, in clinical settings it is always essential to follow up with parents to learn more about the meanings that they are attributing to screening and comprehensive assessment questions and the ways they see their children's behaviors transacting with family values, goals, and the child's broader contextual caregiving environment.

Fourth, the National Institute of Mental Health's new framework for advancing mental health research, the Research Domain Criteria (RDoC), emphasizes the importance of dimensional assessments that characterize problems along a normal to abnormal spectrum and chart the developmental trajectories of functioning over the

course of childhood (Casey et al., 2014; Cuthbert & Insel, 2013; Insel et al., 2010). A core goal of the RDoC initiative is to employ methods that strengthen efforts to link phenotypic manifestations with underlying biological processes (e.g., neural pathways, psychophysiology, genetic risk). Measures that capture discrete, narrowly specified aspects of mood or behavior (e.g., irritability, fear responses) will advance these goals. Such narrowband approaches have proven to be effective in genetics research (Krueger, 1999; Krueger et al., 2002; Tackett et al., 2013; Todd et al., 2005). In our recent work with preschoolers, narrowband dimensional assessments have proven to be effective for identifying linkages with neurocognitive indicators of risk—with evidence that mother-reported difficulty learning from punishment predicted task-based indicators of impairment in learning from punishment (Briggs-Gowan et al., 2013) and observed fear reactions in the Anx-DOS correlated with attention bias to threat, an attentional pattern commonly associated with anxiety (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van, 2007; Briggs-Gowan et al., in press). However, methods for integrating dimensional assessments of social-emotional and behavioral functioning with other methods of characterizing neurobiologic risk will be essential to fully capitalize on these innovative technologies.

Fifth, there is increasing knowledge about the impact of adverse environments and stress on young children (Cicchetti & Toth, 1995; Curtis & Cicchetti, 2011; Finkelhor, Turner, Shattuck, & Hamby, 2013; Finkelhor, Vanderminden, Turner, Shattuck, & Hamby, 2014; Shonkoff & Phillips, 2000; Shonkoff et al., 2012). There is particularly strong evidence that maltreatment and other forms of violence within families are associated with a wide array of social-emotional and behavioral problems in young children. However, we do not know why there is so much heterogeneity in the outcomes associated with apparently similar adverse experiences, ranging from no elevations in social-emotional problems to a heterogeneous set of frank, persistent, impairing, and often comorbid psychiatric disorders. As we become more sophisticated about enriching environments for children with specific social-emotional and behavioral deficits, delays and problems in the forms of early parent-child and child interventions, we need to ensure that the measures we are using are sufficiently sensitive to both maturational and intervention effects.

Given the increased attention to and understanding of the importance of early childhood in laying the foundation for lifelong mental health, it is a very exciting time to enhance the use of evidence-based psychometrically sound,

validated tools. Such tools have been and will continue to be essential for early identification and comprehensive assessment. These also have enabled the scientific community to demonstrate the early problems matter, are persistent and predict functioning later in childhood and drive policy change. While there is still a great deal of work to be done, the extensive attention to measurement development in this developmental period over the past 20 years has enabled a sea of change in understanding of and attention to very young children's mental health needs.

REFERENCES

- Abidin, R. R. (1990). *Parenting Stress Index Short Form-Test manual*. Charlottesville, VA: Pediatric Psychology Press.
- Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs*, *80*(7), 1–37.
- Achenbach, T. M., Edelbrock, C., & Howell, C. T. (1987). Empirically based assessment of the behavioral/emotional problems of 2- and 3-year-old children. *Journal of Abnormal Child Psychology*, *15*, 629–650.
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*(2), 213–232.
- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles*. Burlington, VT: University of Vermont.
- Alvidrez, J. (1999). Ethnic Variations in Mental Health Attitudes and Service Use Among Low-Income African American, Latina, and European American Young Women. *Community Ment Health J*, *35*(6), 515–530. doi: 10.1023/A:1018759201290
- Angold, A., Erkanli, A., Copeland, W., Goodman, R., Fisher, P. W., & Costello, E. J. (2012). Psychiatric diagnostic interviews for children and adolescents: A comparative study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(5), 506–517. doi: 10.1016/j.jaac.2012.02.020
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: Text revision DSM-IV-TR* (Fourth ed.). Washington, DC: American Psychiatric Association.
- Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van, I. M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychological Bulletin*, *133*(1), 1–24. doi: 10.1037/0033-2909.133.1.1
- Barlow, A., Mullany, B., Neault, N., Compton, S., Carter, A., Hastings, R., . . . , Walkup, J. T. (2013). Effect of a paraprofessional home-visiting intervention on American Indian teen mothers' and infants' behavioral risks: A randomized controlled trial. *American Journal of Psychiatry*, *170*(1), 83–93. doi: 10.1176/appi.ajp.2012.12010121
- Bartko, J. J. (1976). On the various intraclass correlation reliability coefficients. *Psychological Bulletin*, *83*, 762–765.
- Ben-Sasson, A., Carter, A. S., & Briggs-Gowan, M. J. (2009). Sensory over-responsivity in elementary school: Prevalence and social-emotional correlates. *Journal of Abnormal Child Psychology*. doi: 10.1007/s10802-008-9295-8
- Ben-Sasson, A., Soto, T. W., Heberle, A. E., Carter, A. S., & Briggs-Gowan, M. J. (2014). Early and concurrent features of ADHD and sensory over-responsivity symptom clusters. *Journal of Attention Disorders*. doi: 10.1177/1087054714543495

- Birmaher, B., Ehmann, M., Axelson, D. A., Goldstein, B. I., Monk, K., Kalas, C., ... Brent, D. A. (2009). Schedule for affective disorders and schizophrenia for school-age children (K-SADS-PL) for the assessment of preschool children—A preliminary psychometric study. *Journal of Psychiatric Research*, *43*(7), 680–686. doi: 10.1016/j.jpsychires.2008.10.003
- Bracha, Z., Perez-Diaz, F., Gerardin, P., Perriot, Y., de la Rocque, F., Flament, M., ... Carter, A. S. (2004). A French adaptation of the Infant–Toddler Social and Emotional Assessment. *Infant Mental Health Journal*, *25*(2), 117–129. doi: 10.1002/imhj.10090
- Bridgett, D. J., Burt, N. M., Laake, L. M., & Oddi, K. B. (2013). Maternal self-regulation, relationship adjustment, and home chaos: Contributions to infant negative emotionality. *Infant Behavior and Development*, *36*(4), 534–547. doi: 10.1016/j.infbeh.2013.04.004
- Briggs-Gowan, M., & Carter, A. S. (2006). *Brief Infant Toddler Social Emotional Assessment (BITSEA)*. San Antonio, TX: Pearson.
- Briggs-Gowan, M., Horowitz, S., & Carter, A. (1998). *The Family Life Impairment Scale*. New Haven, CT: Yale University Press.
- Briggs-Gowan, M. J., & Carter, A. S. (1998). Preliminary acceptability and psychometrics of the Infant-Toddler Social and Emotional Assessment (ITSEA): A new adult-report questionnaire. *Infant Mental Health Journal*, *19*(4), 422–445.
- Briggs-Gowan, M. J., & Carter, A. S. (2008a). Applying the Infant-Toddler Social & Emotional Assessment (ITSEA) and Brief-ITSEA in early intervention. *Infant Mental Health Journal*, *28*, 564–583.
- Briggs-Gowan, M. J., & Carter, A. S. (2008b). Social-emotional screening status in early childhood predicts elementary school outcomes. *Pediatrics*, *121*(5), 957–962. doi: 10.1542/peds.2007-1948
- Briggs-Gowan, M. J., Carter, A. S., Bosson-Heenan, J., Guyer, A. E., & Horwitz, S. M. (2006). Are infant-toddler social-emotional and behavioral problems transient? *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*(7), 849–858. doi: 10.1097/01.chi.0000220849.48650.59
- Briggs-Gowan, M. J., Carter, A. S., Clark, R., Augustyn, M., McCarthy, K. J., & Ford, J. D. (2010). Exposure to potentially traumatic events in early childhood: differential links to emergent psychopathology. *Journal of Child Psychology and Psychiatry*, *51*(10), 1132–1140. doi: 10.1111/j.1469-7610.2010.02256.x
- Briggs-Gowan, M. J., Carter, A. S., & Ford, J. D. (2011). Parsing the effects violence exposure in early childhood: Modeling developmental pathways. *Journal of Pediatric Psychology*, *37*(1), 11–22. doi: 10.1093/jpepsy/jsr063
- Briggs-Gowan, M. J., Carter, A. S., McCarthy, K., Augustyn, M., Caronna, E., & Clark, R. (2013). Clinical validity of a brief measure of early childhood social-emotional/behavioral problems. *Journal of Pediatric Psychology*, *38*(5), 577–587. doi: 10.1093/jpepsy/jst014
- Briggs-Gowan, M. J., Carter, A. S., Skuban, E. M., & Horwitz, S. M. (2001). Prevalence of social-emotional and behavioral problems in a community sample of 1- and 2-year-old children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(7), 811–819.
- Briggs-Gowan, M. J., Ford, J. D., Fraleigh, L., McCarthy, K., & Carter, A. S. (2010). Prevalence of exposure to potentially traumatic events in a healthy birth cohort of very young children in the northeastern United States. *Journal of Traumatic Stress*, *23*(6), 725–733. doi: 10.1002/jts.20593
- Briggs-Gowan, M. J., Horwitz, S. M., Schwab-Stone, M. E., Leventhal, J. M., & Leaf, P. J. (2000). Mental health in pediatric settings: Distribution of disorders and factors related to service use. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(7), 841–849.
- Briggs-Gowan, M. J., Mian, N., Carter, A. S., & Wakschlag, L. S. (2011). *The Anxiety Dimensional Observation Schedule (Anx-DOS), Version 1.0*. Chicago, IL: Northwestern University Press.
- Briggs-Gowan, M. J., Nichols, S. R., Voss, J., Zobel, E., Carter, A. S., McCarthy, K. J., ... Wakschlag, L. S. (2013). Punishment insensitivity and impaired reinforcement learning in preschoolers. *Journal of Child Psychology and Psychiatry*. doi: 10.1111/jcpp.12132
- Briggs-Gowan, M. J., Owens, P. L., Schwab-Stone, M. E., Leventhal, J. M., Leaf, P. J., & Horwitz, S. M. (2003). Persistence of psychiatric disorders in pediatric settings. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(11), 1360–1369. doi: 10.1097/01.CHI.0000084834.67701.8a
- Briggs-Gowan, M. J., Pollak, S. D., Grasso, D., Voss, J., Zobel, E., Wakschlag, L. S., & Pine, D. S. (in press). Attention bias and anxiety in young children exposed to family violence *Journal of Child Psychology and Psychiatry*.
- Broeren, S., & Muris, P. (2008). Psychometric evaluation of two new parent-rating scales for measuring anxiety symptoms in young Dutch children. *Journal of Anxiety Disorders*, *22*, 949–958.
- Brown, G., Marshall, M., Bower, P., Woodham, A., & Waheed, W. (2014). Barriers to recruiting ethnic minorities to mental health research: a systematic review. *International Journal of Methods in Psychiatric Research*, *23*(1), 36–48.
- Bufferd, S. J., Dougherty, L. R., Carlson, G. A., & Klein, D. N. (2011). Parent-reported mental health in preschoolers: Findings using a diagnostic interview. *Comprehensive Psychiatry*, *52*(4), 359–369. doi: 10.1016/j.comppsy.2010.08.006
- Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology & Psychiatry*, *36*, 113–149.
- Carlson, C. I. (1990). Assessing the family context. In C. R. R. W. Kamphaus (Ed.), *Handbook of psychological and educational assessment of children: Personality, behavior, and context* (pp. 546–575). Guilford Press. New York, NY.
- Carter, A. S., & Briggs-Gowan, M. J. (2005). *ITSEA/BITSEA Infant Toddler and Brief Infant Toddler Social Emotional Assessment examiner's manual*. San Antonio, TX: Harcourt Assessment, Inc.
- Carter, A. S., & Briggs-Gowan, M. J. (2006). *ITSEA Infant-Toddler Social and Emotional Assessment examiner's manual*. San Antonio, TX: Harcourt Assessment, Inc.
- Carter, A. S., Briggs-Gowan, M. J., Jones, S. M., & Little, T. D. (2003). The Infant-Toddler Social and Emotional Assessment (ITSEA): Factor structure, reliability, and validity. *Journal of Abnormal Child Psychology*, *31*(5), 495–514.
- Carter, A. S., Little, C., Briggs-Gowan, M. J., & Kogan, N. (1999). The infant-toddler social and emotional assessment (ITSEA): Comparing parent ratings to laboratory observations of task mastery, emotion regulation, coping behaviors and attachment status. *Infant Mental Health Journal*, *20*(4), 375–392.
- Carter, A. S., Marakovitz, S., & Sparrow, S. S. (2006). Comprehensive psychological assessment: A developmental psychopathology approach for clinical and applied research. In D. Cicchetti, & D. J. Cohen (Eds.), *Developmental Psychopathology* (2nd ed., Volume I: Theory and Method, pp. 181–210). New York: John Wiley.
- Carter, A. S., Wagmiller, R. J., Gray, S. A., McCarthy, K. J., Horwitz, S. M., & Briggs-Gowan, M. J. (2010). Prevalence of DSM-IV disorder in a representative, healthy birth cohort at school entry: Sociodemographic risks and social adaptation. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(7), 686–698. doi: 10.1016/j.jaac.2010.03.018
- Casey, B. J., Oliveri, M. E., & Insel, T. (2014). A neurodevelopmental perspective on the research domain criteria (RDoC) framework. *Biological Psychiatry*, *76*(5), 350–353. doi: 10.1016/j.biopsych.2014.01.006

- Cicchetti, D., & Toth, S. L. (1995). A developmental psychopathology perspective on child abuse and neglect. *Journal of the American Academy of Child & Adolescent Psychiatry, 34*(5), 541–565. doi: 10.1097/00004583-199505000-00008
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290. doi: 10.1037/1040-3590.6.4.284
- Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127–137.
- Cicchetti, D. V., Volkmar, F., Klin, A., & Showalter, D. (1995). Diagnosing autism using ICD-10 criteria: A comparison of neural networks and standard multivariate procedures. *Child Neuropsychology, 1*(1), 26–37.
- Clark, R., Tluczek, A., & Gallagher, K. (2004). Assessment of Parent-Child Early Relational Disturbances. In R. DelCarmen-Wiggins & A. S. Carter (Eds.), *Handbook of infant and young child mental health assessment*. Oxford, UK: Oxford University Press.
- Clarke, M. C., Kelleher, I., Clancy, M., & Cannon, M. (2012). Predicting risk and the emergence of schizophrenia. *Psychiatric Clinics of North America, 35*(3), 585–612. doi: 10.1016/j.psc.2012.06.003
- Comer, J. S., Puliafico, A. C., Aschenbrand, S. G., McKnight, K., Robin, J. A., Goldfine, M. E., & Albano, A. M. (2012). A pilot feasibility evaluation of the CALM Program for anxiety disorders in early childhood. *Journal of Anxiety Disorders, 26*(1), 40–49. doi: 10.1016/j.janxdis.2011.08.011
- Compas, B. E. (1987). Coping with stress during childhood and adolescence. *Psychological Bulletin, 101*(3), 393–403.
- Conners, C. K. (2009). *Conners Early Childhood (Conners EC) manual*. North Tonawanda, NY: Multi-Health Systems Inc.
- Conners, C. K., Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): Factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26*(4), 279–291.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.
- Curtis, W. J., & Cicchetti, D. (2011). Affective facial expression processing in young children who have experienced maltreatment during the first year of life: An event-related potential study. *Development and Psychopathology, 23*(2), 373–395. doi: 10.1017/S0954579411000125
- Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med, 11*, 126. doi: 10.1186/1741-7015-11-126
- D'Aprano, A. L., Carapetis, J. R., & Andrews, R. (2011). Trial of a developmental screening tool in remote Australian Aboriginal communities: a cautionary tale. *Journal of Paediatrics and Child Health, 47*(1–2), 12–17. doi: 10.1111/j.1440-1754.2010.01883.x
- Danzig, A. P., Bufferd, S. J., Dougherty, L. R., Carlson, G. A., Olino, T. M., & Klein, D. N. (2013). Longitudinal associations between preschool psychopathology and school-age peer functioning. *Child Psychiatry and Human Development, 44*(5), 621–632. doi: 10.1007/s10578-012-0356-4
- Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Annual research review: Embracing not erasing contextual variability in children's behavior—Theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry, 53*(5), 558–574. doi: 10.1111/j.1469-7610.2012.02537.x
- Disabilities, C. o. C. W. (2001). Developmental surveillance and screening of infants and young children. *Pediatrics, 108*(1), 192–196. doi: 10.1542/peds.108.1.192
- Dougherty, L. R., Smith, V. C., Bufferd, S. J., Stringaris, A., Leibenluft, E., Carlson, G. A., & Klein, D. N. (2013). Preschool irritability: longitudinal associations with psychiatric disorders at age 6 and parental psychopathology. *Journal of the American Academy of Child & Adolescent Psychiatry, 52*(12), 1304–1313. doi: 10.1016/j.jaac.2013.09.007
- Dressler, W. W. (1985). Extended family relationships, social support, and mental health in a southern Black community. *Journal of Health and Social Behavior, 26*(1), 39–48.
- Edwards, S. L., Rapee, R. M., Kennedy, S. J., & Spence, S. H. (2010). The assessment of anxiety symptoms in preschool-aged children: The revised Preschool Anxiety Scale. *Journal of Clinical Child & Adolescent Psychology, 39*(3), 400–409. doi: 10.1080/15374411003691701
- Egger, H. L., & Angold, A. (2004). The Preschool Age Psychiatric Assessment (PAPA): A structured parent interview. In R. DelCarmen-Wiggins & A. Carter (Eds.), *Handbook of infant, toddler, and preschool mental assessment* (pp. 223–243). New York, NY: Oxford University Press.
- Egger, H. L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry, 47*(3–4), 313–337. doi: 10.1111/j.1469-7610.2006.01618.x
- Egger, H. L., Erkanli, A., Keeler, G., Potts, E., Walter, B. K., & Angold, A. (2006). Test-retest reliability of the Preschool Age Psychiatric Assessment (PAPA). *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(5), 538–549. doi: 10.1097/01.chi.0000205705.71194.b8
- Einfeld, S. L., Piccinin, A. M., Mackinnon, A., Hofer, S. M., Taffe, J., Gray, K. M., . . . Tonge, B. J. (2006). Psychopathology in young people with intellectual disability. *JAMA, 296*(16), 1981–1989. doi: 10.1001/jama.296.16.1981
- Ellingson, K. D., Briggs-Gowan, M. J., Carter, A. S., & Horwitz, S. M. (2004). Parent identification of early emerging child behavior problems: Predictors of sharing parental concern with health providers. *Archives of Pediatric and Adolescent Medicine, 158*(8), 766–772. doi: 10.1001/archpedi.158.8.766
- Ellis, R., & Simmons, T. (2014). *Co-resident grandparents and their grandchildren: 2012*. In *Current Population Reports*. U.S. Census Bureau, Washington, DC.
- Emerson, E. (2003). Prevalence of psychiatric disorders in children and adolescents with and without intellectual disability. *Journal of Intellectual Disability Research, 47*, 51–58.
- Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist, 59*(2), 77–92. doi: 10.1037/0003-066X.59.2.77
- Eyberg, S., & Pincus, D. (1999). *Eyberg Child Behavior Inventory (ECBI) & Sutter-Eyberg Student Behavior Inventory-Revised*. Lutz, FL: Psychological Assessment Services.
- Finkelhor, D., Turner, H. A., Shattuck, A., & Hamby, S. L. (2013). Violence, crime, and abuse exposure in a national sample of children and youth: An update. *JAMA Pediatrics, 167*(7), 614–621. doi: 10.1001/jamapediatrics.2013.42
- Finkelhor, D., Vanderminden, J., Turner, H., Shattuck, A., & Hamby, S. (2014). Youth exposure to violence prevention programs in a national sample. *Child Abuse & Neglect*. doi: 10.1016/j.chiabu.2014.01.010
- Fisher, P. W., Shaffer, D., Piacentini, J. C., Lapkin, J., Kafantaris, V., Leonard, H., & Herzog, D. B. (1993). Sensitivity of the Diagnostic Interview Schedule for Children, 2nd ed. (DISC-2.1) for specific diagnoses of children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*(3), 666–673. doi: 10.1097/00004583-199305000-00026
- Flink, I. J., Restrepo, M. H., Blanco, D. P., Ortegón, M. M., Enriquez, C. L., Beirens, T. M., & Raat, H. (2013). Mental health of internally

- displaced preschool children: A cross-sectional study conducted in Bogota, Colombia. *Social Psychiatry and Psychiatric Epidemiology*, 48(6), 917–926. doi: 10.1007/s00127-012-0611-9
- Friday, J. C. (1995). The psychological impact of violence in underserved communities. *Journal of Health Care for the Poor and Underserved*, 6(4), 403–409.
- Fuchs, S., Klein, A. M., Otto, Y., & von Klitzing, K. (2013). Prevalence of emotional and behavioral symptoms and their impact on daily life activities in a community sample of 3 to 5-year-old children. *Child Psychiatry and Human Development*, 44(4), 493–503. doi: 10.1007/s10578-012-0343-9
- Gadomski, A., Wissow, L. S., Slade, E., & Jenkins, P. (2010). Training clinicians in mental health communication skills: Impact on primary care utilization. *Academic Pediatrics*, 10(5), 346–352. doi: 10.1016/j.acap.2010.06.012
- Gaffrey, M. S., Belden, A. C., & Luby, J. L. (2011). The 2-week duration criterion and severity and course of early childhood depression: Implications for nosology. *Journal of Affective Disorders*, 133(3), 537–545. doi: 10.1016/j.jad.2011.04.056
- Gagne, J. R., Van Hulle, C. A., Aksan, N., Essex, M. J., & Goldsmith, H. H. (2011). Deriving childhood temperament measures from emotion-eliciting behavioral episodes: Scale construction and initial validation. *Psychological Assessment*, 23(2), 337–353. doi: 10.1037/a0021746
- Gardner, L. M., Murphy, L., Campbell, J. M., Tylavsky, F., Palmer, F. B., & Graff, J. C. (2013). Screening accuracy for risk of autism spectrum disorder using the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). *Research in Autism Spectrum Disorders*, 7(5), 591–600. doi: 10.1016/j.rasd.2013.01.004
- Gleason, M. M., Zamfirescu, A., Egger, H. L., Nelson, C. A. III, Fox, N. A., & Zeanah, C. H. (2011). Epidemiology of psychiatric disorders in very young children in a Romanian pediatric setting. *European Child & Adolescent Psychiatry*, 20(10), 527–535. doi: 10.1007/s00787-011-0214-0
- Gleason, M. M., Zeanah, C. H., & Dickstein, S. (2010). Recognizing young children in need of mental health assessment: Development and preliminary validity of the early childhood screening assessment. *Infant Mental Health Journal*, 31(3), 335–357. doi: 10.1002/imhj.20259
- Godoy, L., Carter, A. S., Silver, R. B., Dickstein, S., & Seifer, R. (2014). Infants and toddlers left behind: Mental health screening and consultation in primary care. *Journal of Developmental & Behavioral Pediatrics*, 35(5), 334–343. doi: 10.1097/DBP.00000000000004703-201406000-00004 [pii]
- Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4), 400–403. doi: 10.1097/CHI.0b013e3181985068S0890-8567(09)60047-2 [pii]
- Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, 38(8), 1179–1191. doi: 10.1007/s10802-010-9434-x
- Goodman, R., Ford, T., Corbin, T., & Meltzer, H. (2004). Using the Strengths and Difficulties Questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. *European Child & Adolescent Psychiatry*, 13 Suppl 2, II25–31. doi: 10.1007/s00787-004-2005-3
- Gray, S., & Wakschlag, L. (in press). The contribution of developmentally sensitive measurement to assessment of disruptive behavior in young children. In R. DelCarmen-Wiggins & A. S. Carter (Ed.), *Handbook of infant, toddler and preschool mental health assessment* (2nd ed.). Oxford University Press.
- Gray, S. A., Carter, A. S., Briggs-Gowan, M. J., Hill, C., Danis, B., Keenan, K., & Wakschlag, L. S. (2012). Preschool children's observed disruptive behavior: Variations across sex, interactional context, and disruptive psychopathology. *Journal of Clinical Child and Adolescent Psychology*, 41(4), 499–507.
- Gresham, F., & Elliott, S. N. (2008). *Social Skills Improvement System (SSIS) Rating Scales*. Pearson. San Antonio, TX.
- Haley, D. W., & Stansbury, K. (2003). Infant stress and parent responsiveness: Regulation of physiology and behavior during still-face and reunion. *Child Development*, 74(5), 1534–1546.
- Harrison, P. L., & Oakland, T. (2003). *Adaptive Behavior Assessment System, second edition (ABAS-II)*. San Antonio, TX: Psychological Corporation.
- Hayden, L. C., Schiller, M., Dickstein, S., Seifer, R., Sameroff, A. J., Miller, I., ... Rasmussen, S. (1998). Levels of family assessment: I. Family, marital, and parent-child interaction. *Journal of Family Psychology*, 12(1), 7–22.
- Heberle, A. E., Thomas, Y. M., Wagmiller, R. L., Briggs-Gowan, M. J., & Carter, A. S. (2014). The impact of neighborhood, family, and individual risk factors on toddlers' disruptive behavior. *Child Development*, 85(5), 2046–2061. doi: 10.1111/cdev.12251
- Henrichs, J., Rescorla, L., Donkersloot, C., Schenk, J. J., Raat, H., Jaddoe, V. W., ... Tiemeier, H. (2013). Early vocabulary delay and behavioral/emotional problems in early childhood: The generation R study. *Journal of Speech Language and Hearing Research*, 56(2), 553–566. doi: 10.1044/1092-4388(2012)11-0169
- Hensley, M. L., & Spriggs, D. R. (2004). Cancer screening: How good is good enough? *Journal of Clinical Oncology*, 22(20), 4037–4039. doi: 10.1200/JCO.2004.06.129
- Horwitz, S. M., Gary, L. C., Briggs-Gowan, M. J., & Carter, A. S. (2003). Do needs drive services use in young children? *Pediatrics*, 112(6 Pt 1), 1373–1378.
- Horwitz, S. M., Kelleher, K. J., Stein, R. E., Storfer-Isser, A., Youngstrom, E. A., Park, E. R., ... Hoagwood, K. E. (2007). Barriers to the identification and management of psychosocial issues in children and maternal depression. *Pediatrics*, 119(1), e208–218. doi: 10.1542/peds.2005-1997
- Hunsley, J., & Mash, E.J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, 3, 29–51.
- Ialongo, N., Edelsohn, G., Werthamer-Larsson, L., Crockett, L., & Kellam, S. (1993). Are self-reported depressive symptoms in first-grade children developmentally transient phenomena? A further look. *Development and Psychopathology: Special issue: Advances and Challenges in the Study of the Sequelae of Child Maltreatment*, 5, 433–457.
- Ialongo, N., Edelsohn, G., Werthamer-Larsson, L., Crockett, L., & Kellam, S. (1995). The significance of self-reported anxious symptoms in first grade children: Prediction to anxious symptoms and adaptive functioning in fifth grade. *Journal of Child Psychology and Psychiatry*, 36(3), 427–437.
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., ... Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167(7), 748–751. doi: 10.1176/appi.ajp.2010.09091379
- Jianduan, Z., Huishan, W., Shuhua, S., Xiaonan, H., Guoyan, L., Guangli, L., & Junxin, S. (2009). Reliability and validity of standardized Chinese version of Urban Infant-Toddler Social and Emotional Assessment. *Early Human Development*, 85(5), 331–336. doi: 10.1016/j.earlhumdev.2008.12.012

- Kagan, J., Reznick, J.S., Clarke, C., Snidman, N., & Garcia-Coll, C. (1984). Behavioral inhibition to the unfamiliar. *Child Development, 55*(6), 2212–2225.
- Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 Behavioral and Emotional Screening System (BASC-2 BESS)*. San Antonio, TX: Pearson.
- Karabekiroglu, K., Uslu, R., Kapci-Seyitoglu, E. G., Ozbaran, B., Oztop, D. B., Ozel-Ozcan, O., ... Cengel-Kultur, E. S. (2013). A nationwide study of social-emotional problems in young children in Turkey. *Infant Behavior and Development, 36*(1), 162–170. doi: 10.1016/j.infbeh.2012.11.007
- Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among U.S. children: Variation by ethnicity and insurance status. *American Journal of Psychiatry, 159*(9), 1548–1555.
- Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., ... Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children—Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(7), 980–988. doi: 10.1097/00004583-199707000-00021
- Kaufman, J., Birmaher, B., Brent, D., Rao, U., & Ryan, N. (1996). The schedule for affective disorders and schizophrenia for school-age children. Pittsburgh: University of Pittsburgh Medical Center.
- Keenan, K., Shaw, D. S., Walsh, B., Delliquadri, E., & Giovannelli, J. (1997). DSM-III-R disorders in preschool children from low-income families. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(5), 620–627.
- Keenan, K., & Wakschlag, L. S. (2000). More than the terrible twos: The nature and severity of behavior problems in clinic-referred preschool children. *Journal of Abnormal Child Psychology, 28*(1), 33–46.
- Keenan, K., & Wakschlag, L. S. (2002). Can a valid diagnosis of disruptive behavior disorder be made in preschool children? *American Journal of Psychiatry, 159*(3), 351–358.
- Keenan, K., Wakschlag, L. S., Danis, B., Hill, C., Humphries, M., Duax, J., & Donald, R. (2007). Further evidence of the reliability and validity of DSM-IV ODD and CD in preschool children. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*(4), 457–468. doi: 10.1097/CHI.0b013e31803062d300004583-200704000-00007 [pii]
- Kennedy, S. J., Rapee, R. M., & Edwards, S. L. (2009). A selective intervention program for inhibited preschool-aged children of parents with an anxiety disorder: Effects on current anxiety disorders and temperament. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*(6), 602–609.
- Kim-Cohen, J., Arseneault, L., Caspi, A., Tomas, M. P., Taylor, A., & Moffitt, T. E. (2005). Validity of DSM-IV conduct disorder in 4½-5-year-old children: A longitudinal epidemiological study. *American Journal of Psychiatry, 162*(6), 1108–1117. doi: 10.1176/appi.ajp.162.6.1108
- King, T. M., Tandon, S. D., Macias, M. M., Healy, J. A., Duncan, P. M., Swigonski, N. L., ... Lipkin, P. H. (2010). Implementing developmental screening and referrals: Lessons learned from a national project. *Pediatrics, 125*(2), 350–360. doi: 10.1542/peds.2009-0388
- Kleinman, A., Eisenberg, L., & Good, B. (1978). Culture, illness, and care: Clinical lessons from anthropologic and cross-cultural research. *Annals of Internal Medicine, 88*(2), 251–258.
- Kogan, N., & Carter, A. S. (1996). Mother-infant reengagement following the still-face: The role of maternal emotional availability an infant affect regulation. *Infant Behavior and Development, 19*(3), 359–370.
- Kristensen, S., Henriksen, T. B., & Bilenberg, N. (2010). The Child Behavior Checklist for Ages 1.5-5 (CBCL/1½-5): Assessment and analysis of parent- and caregiver-reported problems in a population-based sample of Danish preschool children. *Nordic Journal of Psychiatry, 64*(3), 203–209. doi: 10.3109/08039480903456595
- Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry, 56*(10), 921–926.
- Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior, and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology, 111*(3), 411–424.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.
- Laughlin, L. (2013). Who's minding the kids? Child care arrangements: Spring 2011. In *Current Population Reports* (pp. 70–135). Washington, DC: U.S. Census Bureau.
- Lavigne, J. V., Arend, R., Rosenbaum, D., Binns, H. J., Christoffel, K. K., & Gibbons, R. D. (1998). Psychiatric disorders with onset in the preschool years: II. Correlates and predictors of stable case status. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*(12), 1255–1261.
- Lavigne, J. V., Gibbons, R. D., Christoffel, K. K., Arend, R., Rosenbaum, D., Binns, H., ... Isaacs, C. (1996). Prevalence rates and correlates of psychiatric disorders among preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*(2), 204–214.
- Lavigne, J. V., Lebailly, S. A., Hopkins, J., Gouze, K. R., & Binns, H. J. (2009). The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. *Journal of Clinical Child and Adolescent Psychology, 38*(3), 315–328. doi: 10.1080/15374410902851382
- Liu, J., Cheng, H., & Leung, P. W. (2011). The application of the preschool Child Behavior Checklist and the caregiver-teacher report form to Mainland Chinese children: Syndrome structure, gender differences, country effects, and inter-informant agreement. *Journal of Abnormal Child Psychology, 39*(2), 251–264. doi: 10.1007/s10802-010-9452-8
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., ... Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism & Developmental Disorders, 30*(3), 205–223.
- Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders, 19*(2), 185–212.
- Lowell, D. I., Carter, A. S., Godoy, L., Paulicin, B., & Briggs-Gowan, M. J. (2011). A randomized controlled trial of Child FIRST: A comprehensive home-based intervention translating research into early childhood practice. *Child Development, 82*(1), 193–208. doi: 10.1111/j.1467-8624.2010.01550.x
- Luby, J., Belden, A., Sullivan, J., Hayen, R., McCadney, A., & Spitznagel, E. (2009). Shame and guilt in preschool depression: Evidence for elevations in self-conscious emotions in depression as early as age 3. *Journal of Child Psychology and Psychiatry, 50*(9), 1156–1166. doi: 10.1111/j.1469-7610.2009.02077.x
- Luby, J. L., Belden, A. C., Pautsch, J., Si, X., & Spitznagel, E. (2009). The clinical significance of preschool depression: Impairment in functioning and clinical markers of the disorder. *Journal of Affective Disorders, 112*(1–3), 111–119. doi: 10.1016/j.jad.2008.03.026
- Luby, J. L., Heffelfinger, A. K., Mrakotsky, C., Hessler, M. J., Brown, K. M., & Hildebrand, T. (2002). Preschool major depressive disorder: preliminary validation for developmentally modified DSM-IV criteria. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(8), 928–937.

- Luby, J. L., & Navsaria, N. (2010). Pediatric bipolar disorder: Evidence for prodromal states and early markers. *Journal of Child Psychology and Psychiatry*, 51(4), 459–471. doi: 10.1111/j.1469-7610.2010.02210.x
- Luby, J. L., Si, X., Belden, A. C., Tandon, M., & Spitznagel, E. (2009). Preschool depression: Homotypic continuity and course over 24 months. *Archives of General Psychiatry*, 66(8), 897–905. doi: 10.1001/archgenpsychiatry.2009.97
- Lucey, P., Cohn, J. F., Matthews, I., Lucey, S., Sridharan, S., Howlett, J., & Prkachin, K. M. (2011). Automatically detecting pain in video through facial action units. *IEEE Transaction Systems Man Cybernetics B Cybernetics*, 41(3), 664–674. doi: 10.1109/TSMCB.2010.2082525
- Luster, T., Rhoades, K., & Haas, B. (1989). The relation between parental values and parenting behavior: A test of the Kohn hypothesis. *Journal of Marriage and the Family*, 139–147.
- Martinez-Torteya, C., Dayton, C. J., Beeghly, M., Seng, J. S., McGinnis, E., Broderick, A., . . . Muzik, M. (2014). Maternal parenting predicts infant biobehavioral regulation among women with a history of childhood maltreatment. *Development and Psychopathology*, 26(2), 379–392. doi: 10.1017/S0954579414000017
- Martini, D. R., Strayhorn, J. M., & Puig-Antich, J. (1990). A symptom self-report measure for preschool children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 29(4), 594–600. doi: 10.1097/00004583-199007000-00013
- Mathiesen, K. S., & Sanson, A. (2000). Dimensions of early childhood behavior problems: Stability and predictors of change from 18 to 30 months. *Journal of Abnormal Child Psychology*, 28(1), 15–31.
- McCue Horwitz, S., Hurlburt, M. S., Heneghan, A., Zhang, J., Rolls-Reutz, J., Fisher, E., . . . Stein, R. E. (2012). Mental health problems in young children investigated by U.S. child welfare agencies. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(6), 572–581. doi: 10.1016/j.jaac.2012.03.006
- McHale, J. P., & Coates, E. E. (2014). Observed coparenting and triadic dynamics in African American fragile families at 3 months' postpartum. *Infant Mental Health Journal*, 35(5), 435–451.
- McHale, J. P., & Fivaz-Depeursinge, E. (1999). Understanding triadic and family group interactions during infancy and toddlerhood. *Clinical Child and Family Psychology Review*, 2(2), 107–127.
- McLaughlin, B., Gesi Blanchard, A., & Osanai, Y. (1995). Assessing language development in bilingual preschool children, NCBE Program Information Guide Series No. 22. National Clearinghouse for Bilingual Education. Washington, D.C.
- Measelle, J. R., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (1998). Assessing young children's views of their academic, social, and emotional lives: An evaluation of the self-perception scales of the Berkeley Puppet Interview. *Child Development*, 69(6), 1556–1576.
- Merrell, K. W. (2002). *Preschool and Kindergarten Behavior Scales (PKBS-2)*, second edition. Austin, TX: Pro-Ed, Inc.
- Mian, N., Carter, A. S., Pine, D. S., Wakschlag, L., & Briggs-Gowan, M. J. (in press). Development of a novel observational measure for anxiety in young children: The anxiety dimensional observation scale. *Journal of Child Psychology and Psychiatry*. doi: 10.1111/jcpp.12407
- Mian, N. D., Godoy, L., Briggs-Gowan, M. J., & Carter, A. S. (2012). Patterns of anxiety symptoms in toddlers and preschool-age children: Evidence of early differentiation. *Journal of Anxiety Disorders*, 26(1), 102–110. doi: 10.1016/j.janxdis.2011.09.006
- Milgrom, J., Mendelsohn, J., & Gemmill, A. W. (2011). Does postnatal depression screening work? Throwing out the bathwater, keeping the baby. *Journal of Affective Disorders*, 132(3), 301–310. doi: 10.1016/j.jad.2010.09.031
- Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities*, 41(5), 417–436.
- Mouton-Simien, P., McCain, A. P., & Kelley, M. L. (1997). The development of the Toddler Behavior Screening Inventory. *Journal of Abnormal Child Psychology*, 25(1), 59–64.
- Neitzel, C., & Stright, A. D. (2003). Mothers' scaffolding of children's problem solving: Establishing a foundation of academic self-regulatory competence. *Journal of Family Psychology*, 17(1), 147–159.
- Nichols, S., Briggs-Gowan, M. J., Estabrook, R., Burns, J., Kestler, J., Berman, G., . . . Wakschlag, L. S. (in press). Punishment insensitivity in early childhood: A developmental, dimensional approach. *Journal of Abnormal Child Psychology*.
- Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill Publishing Co.
- O'Neill, S., Schneiderman, R. L., Rajendran, K., Marks, D. J., & Halperin, J. M. (2014). Reliable ratings or reading tea leaves: Can parent, teacher, and clinician behavioral ratings of preschoolers predict ADHD at age six? *Journal of Abnormal Child Psychology*, 42(4), 623–634. doi: 10.1007/s10802-013-9802-4
- Otto, Y., Andreas, A., von Klitzing, K., Fuchs, S., & Klein, A. M. (2014). [Sadness, worries and fears: depression and anxiety disorders in preschool age—Results of relevance, symptoms and impairment]. *Prax Kinderpsychol Kinderpsychiatr*, 63(3), 154–176.
- Policy statement—The future of pediatrics: Mental health competencies for pediatric primary care. (2009). *Pediatrics*, 124(1), 410–421. doi: 10.1542/peds.2009-1061
- Qi, C. H., & Kaiser, A. P. (2003). Behavior problems in preschool children from low-income families: Review of the literature. *Topics in Early Childhood Special Education*, 23(4), 188–216.
- Rapee, R. M., Kennedy, S., Ingram, M., Edwards, S., & Sweeney, L. (2005). Prevention and early intervention of anxiety disorders in inhibited preschool children. *Journal of Consulting and Clinical Psychology*, 73(3), 488–497. doi: 10.1037/0022-006X.73.3.488
- Reynolds, C. R., & Kamphaus, R. W. (2004). *BASC-2 Behavior Assessment System for Children: Second edition manual (BASC-2)*. Minneapolis, MN: NCS Pearson.
- Reznick, J. S., Kagan, J., Snidman, N., Gersten, M., Baak, K., & Rosenberg, A. (1986). Inhibited and uninhibited children: A follow-up study. *Child Development*, 57(3), 660–680.
- Richman, N., & Graham, P. J. (1971). A behavioural screening questionnaire for use with three-year-old children. Preliminary findings. *Journal of Child Psychology and Psychiatry*, 12, 5–33.
- Robins, D. L., Casagrande, K., Barton, M., Chen, C. M., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for Autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*, 133(1), 37–45. doi: 10.1542/peds.2013-1813
- Ross, G., & Weinberg, S. (2006). Is there a relationship between language delays and behavior and socialization problems in toddlers? *Journal of Early Childhood and Infant Psychology*, 2, 2101–2116.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328), 918–924.
- Schechter, D. S., & Willheim, E. (2009). The effects of violent experiences on infants and young children. In C. Zeanah (Ed.), *Handbook of infant mental health* (third edition, pp. 197–213). New York, NY: Guilford Press.
- Scheeringa, M. S. (2004). *Diagnostic Infant and Preschool Assessment (DIPA) manual*. Institute of Infant and Early Childhood Mental Health: Section of Child and Adolescent Psychiatry, Department of Psychiatry and Neurology. Tulane University, New Orleans, LA.

- Scheeringa, M. S. (2007). A research agenda for posttraumatic stress disorder in infants, toddlers, and preschool children. In W. E. Narrow, M. B. First, P. J. Sirovatka, & D. A. Regier (Eds.), *Age and gender considerations in psychiatric diagnosis* (pp. 151–162). Arlington, VA: American Psychiatric Association.
- Scheeringa, M. S. (2008). Developmental considerations for diagnosing PTSD and acute stress disorder in preschool and school-age children. *American Journal of Psychiatry*, *165*(10), 1237–1239. doi: 10.1176/appi.ajp.2008.08070974
- Scheeringa, M. S., & Haslett, N. (2010). The reliability and criterion validity of the Diagnostic Infant and Preschool Assessment: A new diagnostic instrument for young children. *Child Psychiatry and Human Development*, *41*(3), 299–312. doi: 10.1007/s10578-009-0169-2
- Scheeringa, M. S., Myers, L., Putnam, F. W., & Zeanah, C. H. (2012). Diagnosing PTSD in early childhood: An empirical assessment of four approaches. *Journal of Traumatic Stress*, *25*(4), 359–367. doi: 10.1002/jts.21723
- Scheeringa, M. S., Peebles, C. D., Cook, C. A., & Zeanah, C. H. (2001). Toward establishing procedural, criterion, and discriminant validity for PTSD in early childhood. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(1), 52–60.
- Scheeringa, M. S., & Zeanah, C. H. (2001). A relational perspective on PTSD in early childhood. *Journal of Traumatic Stress*, *14*(4), 799–815.
- Schonwald, A., Huntington, N., Chan, E., Risko, W., & Bridgemohan, C. (2009). Routine developmental screening implemented in urban primary care settings: More evidence of feasibility and effectiveness. *Pediatrics*, *123*(2), 660–668. doi: 10.1542/peds.2007-2798
- Seifer, R. (2000). Temperament and goodness of fit: Implications for developmental psychopathology. In A. J. Sameroff, M. Lewis, & S. M. Miller (eds), *Handbook of developmental psychopathology* (2nd edition, pp. 257–276). New York: Plenum Press
- Shamir-Essakow, G., Ungerer, J. A., & Rapee, R. M. (2005). Attachment, behavioral inhibition, and anxiety in preschool children. *Journal of Abnormal Child Psychology*, *33*(2), 131–143.
- Shaw, D. S., Lacourse, E., & Nagin, D. (2005). Developmental trajectories of conduct problems and hyperactivity from ages 2 to 10. *Journal of Psychology and Psychiatry*, *46*(9), 931–942.
- Sheldrick, R. C., Henson, B. S., Merchant, S., Neger, E. N., Murphy, J. M., & Perrin, E. C. (2012). The Preschool Pediatric Symptom Checklist (PPSC): Development and initial validation of a new social/emotional screening instrument. *Academic Pediatrics*, *12*(5), 456–467. doi: 10.1016/j.acap.2012.06.008S1876-2859(12)00178-7
- Sheldrick, R. C., Henson, B. S., Neger, E. N., Merchant, S., Murphy, J. M., & Perrin, E. C. (2013). The baby pediatric symptom checklist: Development and initial validation of a new social/emotional screening instrument for very young children. *Academic Pediatrics*, *13*(1), 72–80. doi: 10.1016/j.acap.2012.08.003S1876-2859(12)00202-1
- Sheldrick, R. C., Merchant, S., & Perrin, E. C. (2011). Identification of developmental-behavioral problems in primary care: A systematic review. *Pediatrics*, *128*(2), 356–363. doi: 10.1542/peds.2010-3261
- Sheldrick, R. C., Neger, E. N., & Perrin, E. C. (2012). Concerns about development, behavior, and learning among parents seeking pediatric care. *Journal of Developmental and Behavioral Pediatrics*, *33*(2), 156–160. doi: 10.1097/DBP.0b013e3182420f4a
- Sheldrick, R. C., & Perrin, E. C. (2013). Evidence-based milestones for surveillance of cognitive, language, and motor development. *Academic Pediatrics*, *13*(6), 577–586. doi: 10.1016/j.acap.2013.07.001S1876-2859(13)00202-7
- Shonkoff, J., & Phillips, D. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.
- Shonkoff, J. P., Garner, A. S., & The Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood, Adoption, and Dependent, Care, and Section on Developmental and Behavioral Pediatrics. (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics*, *129*(1), e232–246. doi: 10.1542/peds.2011-2663
- Silverman, W. K., Saavedra, L. M., & Pina, A. A. (2001). Test–retest reliability of anxiety symptoms and diagnoses with the Anxiety Disorders Interview Schedule for DSM-IV: Child and parent versions. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(8), 937–944. doi: 10.1097/00004583-200108000-00016
- Smith, N. J., Sheldrick, R. C., & Perrin, E. C. (2013). An abbreviated screening instrument for autism spectrum disorders. *Infant Mental Health Journal*, *34*(2), 149–155. doi: 10.1002/imhj.21356
- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales: Second edition (Vineland-II), Survey Interview Form/Caregiver Rating Form*. Livonia, MN: Pearson Assessment.
- Speltz, M. L., McClellan, J., DeKlyen, M., & Jones, K. (1999). Preschool boys with oppositional defiant disorder: Clinical presentation and diagnostic change. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*(7), 838–845. doi: 10.1097/00004583-199907000-00013
- Spence, S. H., Najman, J. M., Bor, W., O’Callaghan, M. J., & Williams, G. M. (2002). Maternal anxiety and depression, poverty and marital relationship factors during early childhood as predictors of anxiety and depressive symptoms in adolescence. *Journal of Child Psychology and Psychiatry*, *43*(4), 457–469.
- Spence, S. H., Rapee, R., McDonald, C., & Ingram, M. (2001). The structure of anxiety symptoms among preschoolers. *Behavioral Research and Therapy*, *39*(11), 1293–1316.
- Spittle, A. J., Anderson, P. J., Lee, K. J., Ferretti, C., Eeles, A., Orton, J., . . . Doyle, L. W. (2010). Preventive care at home for very preterm infants improves infant and caregiver outcomes at 2 years. *Pediatrics*, *126*(1), 171–178.
- Squires, J., Bricker, D., & Twombly, E. (2002). *Ages and Stages Questionnaire: Social-Emotional: A parent-completed, child-monitoring system for social-emotional behaviors*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Stalets, M. M., & Luby, J. L. (2006). Preschool depression. *Child and Adolescent Psychiatry Clinics of North America*, *15*(4), 899–917, viii–ix. doi: 10.1016/j.chc.2006.05.011
- Ştefan, C. A., & Miclea, M. (2013). Effects of a multifocused prevention program on preschool children’s competencies and behavior problems. *Psychology in the Schools*, *50*(4), 382–402. doi: 10.1002/pits.21683
- Sterba, S., Egger, H. L., & Angold, A. (2007). Diagnostic specificity and nonspecificity in the dimensions of preschool psychopathology. *Journal of Child Psychology and Psychiatry*, *48*(10), 1005–1013. doi: 10.1111/j.1469-7610.2007.01770.x
- Stoneman, Z., Brody, G. H., Churchill, S. L., & Winn, L. L. (1999). Effects of residential instability on Head Start children and their relationships with older siblings: Influences of child emotionality and conflict between family caregivers. *Child Development*, *70*(5), 1246–1262.
- Tackett, J. L., Lahey, B. B., van Hulle, C., Waldman, I., Krueger, R. F., & Rathouz, P. J. (2013). Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence. *Journal of Abnormal Psychology*, *122*(4), 1142–1153. doi: 10.1037/a0034151
- Thaler, N. S., Allen, D. N., Scott, J. G. (2014). Neuropsychological assessment of Asian American children and adolescents. 407–425.
- Thomas, J. M., & Guskin, K. A. (2001). Disruptive behavior in young children: What does it mean? *Journal of the American Academy of*

- Child & Adolescent Psychiatry*, 40(1), 44–51. doi: 10.1097/00004583-200101000-00015
- Todd, R. D., Huang, H., Smalley, S. L., Nelson, S. F., Willcutt, E. G., Pennington, B. F., . . . Neuman, R. J. (2005). Collaborative analysis of DRD4 and DAT genotypes in population-defined ADHD subtypes. *Journal of Child Psychology & Psychiatry*, 46(10), 1067–1073.
- Tronick, E. Z. (1989). Emotions and emotional communication in infants. *American Psychologist*, 44(2), 112–119.
- van Widenfelt, B. M., Treffers, P. D. A., de Beurs, E., Siebelink, B. M., & Koudijs, E. (2005). Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clinical Child and Family Psychology Review*, 8(2), 135–147.
- Visser, J. C., Smeekens, S., Rommelse, N., Verkes, R. J., Van der Gaag, R. J., & Buitelaar, J. K. (2010). Assessment of psychopathology in 2- to 5-year-olds: Applying the Infant-Toddler Social Emotional Assessment. *Infant Mental Health Journal*, 31(6), 611–629.
- Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development*, 75(1), 296–312.
- Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.
- Wakschlag, L. S., Briggs-Gowan, M. J., Carter, A. S., Hill, C., Danis, B., Keenan, K., . . . Leventhal, B. L. (2007). A developmental framework for distinguishing disruptive behavior from normative misbehavior in preschool children. *Journal of Child Psychology and Psychiatry*, 48(10), 976–987. doi: 10.1111/j.1469-7610.2007.01786.x
- Wakschlag, L. S., Briggs-Gowan, M. J., Choi, S. W., Nichols, S. R., Kestler, J., Burns, J. L., . . . Henry, D. (2014). Advancing a multidimensional, developmental spectrum approach to preschool disruptive behavior. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(1), 82–96 e83. doi: 10.1016/j.jaac.2013.10.011
- Wakschlag, L. S., Briggs-Gowan, M. J., Hill, C., Danis, B., Leventhal, B. L., Keenan, K., . . . Carter, A. S. (2008). Observational Assessment of Preschool Disruptive Behavior, Part II: Validity of the Disruptive Behavior Diagnostic Observation Schedule (DB-DOS). *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(6), 632–641. doi: 10.1097/CHI.0b013e31816c5c10
- Wakschlag, L. S., Choi, S. W., Carter, A. S., Hullsiek, H., Burns, J., McCarthy, K., . . . Briggs-Gowan, M. J. (2012). Defining the developmental parameters of temper loss in early childhood: Implications for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 53(11), 1099–1108. doi: 10.1111/j.1469-7610.2012.02595.x
- Wakschlag, L. S., Estabrook, R., Henry, D., Burns, J. L., Perlman, S. B., Pine, D. S., . . . Briggs-Gowan, M. J. (in press). Clinical application of a dimensional approach: The normal:abnormal spectrum of early irritability. *Journal of the American Academy of Child & Adolescent Psychiatry*.
- Wakschlag, L. S., Henry, D. B., Tolan, P., Carter, A. S., Burns, J., & Briggs-Gowan, M. J. (2012). Putting theory to the test: Modeling a multidimensional, developmentally-based approach to preschool disruptive behavior. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(6), 593–604.
- Wakschlag, L. S., Hill, C., Carter, A. S., Danis, B., Egger, H. L., Keenan, K., . . . Briggs-Gowan, M. J. (2008). Observational assessment of preschool disruptive behavior, part I: reliability of the Disruptive Behavior Diagnostic Observation Schedule (DB-DOS). *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(6), 622–631. doi: 10.1097/CHI.0b013e31816c5bdb
- Wakschlag, L. S., Tolan, P. H., & Leventhal, B. L. (2010). Research review: “Ain’t misbehavin’”: Towards a developmentally-specified nosology for preschool disruptive behavior. *Journal of Child Psychology and Psychiatry*, 51(1), 3–22.
- Wang, M., Zhou, J. (2014). Anxiety disorder symptoms in Chinese preschool children. *Child Psychiatry & Human Development*, 46(1): 158–66.
- Warren, S. L., Emde, R. N., & Sroufe, L. A. (2000). Internal representations: predicting anxiety from children's play narratives. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39(1), 100–107. doi: 10.1097/00004583-200001000-00022
- Warren, S. L., Oppenheim, D., & Emde, R. N. (1996). Can emotions and themes in children's play predict behavior problems? *Journal of the American Academy of Child & Adolescent Psychiatry*, 35(10), 1331–1337. doi: 10.1097/00004583-199610000-00020
- Weitzman, C., Edmonds, D., Davagnino, J., & Briggs-Gowan, M. J. (2014). Young child socioemotional/behavioral problems and cumulative psychosocial risk. *Infant Mental Health Journal*, 35(1), 1–9. doi: 10.1002/imhj.21421
- Wichstrom, L., Berg-Nielsen, T. S., Angold, A., Egger, H. L., Solheim, E., & Sveen, T. H. (2012). Prevalence of psychiatric disorders in preschoolers. *Journal of Child Psychology and Psychiatry*, 53(6), 695–705. doi: 10.1111/j.1469-7610.2011.02514.x
- Wilson, M. N., & Tolson, T. F. (1990). Familial support in the Black community. *Journal of Clinical Child Psychology*, 19(4), 347–355.
- World Health Organization. (2001). *ICF: International Classification of Functioning, Disability, and Health*. Geneva, Switzerland: World Health Organization.
- Wu, Y. T., Chen, W. J., Hsieh, W. S., Chen, P. C., Liao, H. F., Su, Y. N., & Jeng, S. F. (2012). Maternal-reported behavioral and emotional problems in Taiwanese preschool children. *Research in Developmental Disabilities*, 33(3), 866–873. doi: 10.1016/j.ridd.2011.11.018
- Yasui, M., & Dishion, T. J. (2007). The ethnic context of child and adolescent problem behavior: Implications for child and family interventions. *Clinical Child and Family Psychology Review*, 10(2), 137–179. doi: 10.1007/s10567-007-0021-9
- Yu, J. W., Buka, S. L., McCormick, M. C., Fitzmaurice, G. M., & Indurkha, A. (2006). Behavioral problems and the effects of early intervention on eight-year-old children with learning disabilities. *Maternal and Child Health Journal*, 10(4), 329–338. doi: 10.1007/s10995-005-0066-7