Evidence-based Assessment in Pediatric Psychology: Family Measures

Melissa A. Alderfer,1,2 PhD, Barbara H. Fiese,3 PhD, Jeffrey I. Gold,4 PhD, J. J. Cutuli,1,9 BA, Grayson N. Holmbeck,5 PhD, Lutz Goldbeck,6 PhD, Christine T. Chambers,7 PhD, Mona Abad,5 PhD, Dante Spetter,8 PhD, and Joän Patterson,9 PhD
1 The Children’s Hospital of Philadelphia, 2 University of Pennsylvania School of Medicine, 3 Syracuse University, 4 Childrens Hospital Los Angeles, Keck School of Medicine, University of Southern California, 5 Loyola University Chicago, 6 University Hospital Ulm, 7 Dalhousie University and IWK Health Centre, 8 Harvard University, and 9 University of Minnesota

Objective To provide a review of the evidence base of family measures relevant to pediatric psychology.

Method Twenty-nine family measures were selected based upon endorsement by Division 54 listserv members, expert judgment, and literature review. Spanning observational and self-report methods, the measures fell into three broad assessment categories: Family functioning, Dyadic family relationships, and Family functioning in the context of childhood chronic health conditions. Measures were categorized as: “Well-established”, “Approaching well-established”, or “Promising.”

Results Nineteen measures met “well-established” criteria and the remaining ten were “approaching well-established.” “Well-established” measures were documented for each of the broad assessment categories named above.

Conclusions Many measures deemed “well-established” in the general population are proving to be reliable and useful in pediatric samples. More evidence of the validity of family measures is needed in this context. This review should prove helpful to clinicians and researchers as they strive to make evidence-based decisions regarding family measures.

Key words assessment; evidence-based; family measurement; pediatric psychology.
functioning well, roles are clear, communication is open and straightforward, and affect is well regulated. In contrast, when a family is functioning poorly, it may, for example, respond to stress by becoming disorganized and chaotic with unfocused communication patterns and emotional dysregulation. Alternatively, it may become over-controlled, with increased rigidity of roles, inadequate communication, and a restricted range of affect.

General measures of family functioning attempt to assess these important aspects of families. Most measures assess family organization (e.g., roles, leadership, and alliance formation), cohesion (e.g., involvement and closeness), communication (e.g., clarity of expression and directness), affective environment (e.g., expression of feelings and conflict), and problem solving ability (e.g., goal-directed negotiation and task accomplishment) to capture the most elemental components of the operations of the family. Many measures also include a single index of general functioning that reflects overall adaptation. Using a broad-based measure of general functioning may be appropriate or choosing a specific domain of functioning may be indicated depending upon the specific research or clinical question posed. Multiple perspectives within one family (e.g., self-report by parents and children) may also be desired.

It is also important to decide which components of the family system to assess. The family system is more than the sum of its parts, but dyadic subsystems within families (e.g., partner–partner, parent–child, and sibling–sibling relationships) may be of specific interest. Dyadic measurement tools focus on relationship issues such as warmth, attachment, conflict, and competition as opposed to more global family functioning constructs. For questions such as how marital conflict may compromise treatment regimen adherence, such measures are most useful.

A third consideration is specificity of the assessment device. Instead of measuring global family functioning, a pediatric psychologist may be interested in assessing the functioning of the family within the context of the medical condition they are facing. For example, illness-related communication, or parenting stress related to a child’s disability may be of interest. Measures assessing specific constructs such as these are being developed within the field because of their inherent worth both clinically and empirically.

Finally, the method of assessment is an important consideration. Although not unique to family assessment, there are typically modest associations between self-assessments and observer ratings of family functioning. Often referred to as the insider versus outsider view of the family (Fiese & Spagnola, 2006), such discrepancies make decisions regarding assessment method important. Self-report questionnaires, coded observed interactions, and interviews each provide a different window into family functioning, and decisions must be made as to which one or combination is most suitable for one’s purpose.

In an effort to help practitioners and researchers make such decisions, we have reviewed the available psychometric data for 29 family measures and classified them in regard to the quality of their evidence base. For the purpose of this review, we use the term “family” to refer to the entire family unit as well as its subsystems, and included measures of general family functioning and dyadic familial relationships. Our review complements similar recent efforts reported in books (Kerig & Lindahl, 2001; Walsh, 2003) and review papers (DeCato, Donohue, Azrin, Teichner, & Crum, 2002), but is distinct in that it focuses on the use of family measures within pediatric populations. Upon reviewing the available literature, we categorize each measure as “well-established,” “approaching well-established,” or “promising” using Cohen and colleagues’ (2008) criteria.

Method

Work Group Formation

The Work Group Chair (M.A.A.) was invited by the APA Division 54 Assessment Task Force to assemble the group and organize the review. A call for Work Group members was distributed through the Division 54 listserv and from among the respondents seven members were selected (B.F., J.G., G.H., L.G., C.C., D.S., and J.P.). These individuals have significant experience with family measures ranging from repeated use of existing measures to development of new measures. Their individual clinical and research areas broadly represent the field of pediatric psychology and they were drawn from across the nation with two international representatives.

Measure Selection

Measures developed or used within pediatric populations or with an extensive literature base in the general population with promise for use in pediatric populations were selected for review. Results from the Division 54 [Society of Pediatric Psychology (SPP)] listserv survey, described in detail in the Introduction section to this series (Cohen et al., 2008), our own experience,
and a brief review of relevant literature (e.g., papers published in the Journal of Pediatric Psychology) were used to identify family measures in use in the field of pediatric psychology. A total of 44 family-related measures were listed on the Division 54 survey. Twenty-seven of these measures assessed general family functioning, dyadic familial relationships, or families in the context of childhood chronic medical conditions. Twenty-two of them were selected for review. They included self-report questionnaires endorsed by seven or more respondents, observational coding systems used by two or more respondents, and measures specific to pediatric psychology endorsed by at least one respondent, given their inherent relevance to our readership. An additional eight measures were identified and added to the review based upon survey respondents’ recommendations, workgroup members’ experiences, and review of the literature. The list of measures included in the review was not intended to be exhaustive or comprehensive, but rather targeted to include measures relevant to or potentially useful to the field of pediatric psychology. One observational measure of general family functioning (The Family Interaction Coding System; Patterson, Ray, Shaw, & Cobb, 1969), endorsed by 11 respondents in the survey was removed during the review process as the manual is currently out of print and unavailable (G. R. Patterson, personal communication, June 2006). In total, 29 measures were reviewed: 13 assessing general family functioning, eight assessing familial relationships, and eight assessing families in the context of a chronic medical condition. Material for the review was identified through electronic databases (PsycInfo, Medline, and Health and Psychosocial Instruments), reference lists, and contact with the measures’ authors. Material published prior to January 1, 2007 was included in the reviews.

Assessment Criteria

The SPP Assessment Task Force developed specific criteria to be used in this review of assessment measures as detailed in the introductory paper to this series (Cohen et al., in press). They debated criteria and came to consensus regarding a three-tiered hierarchical classification system paralleling those of Chambless and Ollendick (2001) used to evaluate psychological interventions. Work Groups, assigned subsets of the measures, were charged with describing, critiquing, and finally classifying the measures into these “evidence-based categories.”

“Well-established” assessments fulfill the following criteria: (a) they have been presented in at least two peer-reviewed articles by different investigatory teams; (b) sufficient details have been provided to allow for critical evaluation and replication (e.g., a manual is available); and (c) detailed (e.g., statistical) information has been presented in at least one peer-reviewed article indicating good reliability and validity. We defined “good” psychometric properties as follows: internal consistency (coefficient α) ≥ .70 (Nunnally, 1978), test–retest reliability consistent with the purported stability of the construct, inter-rater reliability (α or ICC) ≥ .70 and/or inter-rater agreement (κ) ≥ .61 (Landis & Koch, 1977), and at least two forms of evidence of concurrent/predictive or convergent validity. For classification purposes, we only considered psychometric data reported in empirical, peer-reviewed manuscripts, though our reviews include data from nonpeer reviewed manuals, chapters, and peer-reviewed reviews. When multiple peer-reviewed manuscripts provided psychometric data for a given instrument, the preponderance of the evidence needed to be consistent with our definition of “good” for a measure to qualify as “well-established.” It should be noted that these criteria only set a lower boundary and therefore, there may be great variability in the extent of the evidence base for measures within this category.

Measures qualifying for the “approaching well-established” category fulfill the following criteria: (a) presented in at least two peer-reviewed articles, possibly by the same investigatory team; (b) sufficient detail provided to allow for critical evaluation and replication (e.g., a manual is available); and (c) reliability and validity is moderate at best or presented in vague (i.e., nonnumeric) terms. Here, it should be noted that even if a measure has “good” psychometric properties, if it is only being used by a single investigatory team, it is considered “approaching well-established.” “Promising assessments” have only been presented in one peer-reviewed article, but fulfill criteria (b) and (c) of “approaching well-established” measures.

1A total of 13 self-report measures assessing general family functioning and dyadic relationships within families were included in the survey. All but two of these measures were being used by seven or more survey respondents. The remaining two measures were endorsed by two or fewer respondents. Because this seemed like a natural break in the available data, these two measures were not included in the review.

2A total of nine observational measures of general family functioning and dyadic relationships were included in the survey. One measure was endorsed by no respondents and one was endorsed by one respondent. After reviewing the available literature on these two measures, we decided by consensus to eliminate them from review due to their limited use and evidence base.
To review and classify the 29 identified family assessment measures, our work group was divided into subgroups. Each subgroup provided preliminary reviews and classification of a subset of the measures. These preliminary reviews were expanded as additional information was gathered and the classifications were re-evaluated. Final classification was based upon consensus of the work group members.

**Results**

Complete and comprehensive reviews of the 29 identified family assessment measures: (a) identify the key source and measurement development references; (b) describe the purpose, format, scoring, and applicable age ranges; (c) summarize available psychometric information; (d) evaluate clinical and research utility; (e) discuss applicability with linguistic minorities as well as racial and ethnic populations; and (f) classify each measure according to the criteria outlined earlier. Information from these comprehensive reviews is summarized in the text and accompanying tables. Reviews with all relevant references are available upon request.

Supplementary Table S1 summarizes the psychometric properties of the reviewed instruments. Internal consistency, test–retest, and inter-rater reliability are provided. For observational methods, intraclass correlations (ICCs) or kappas (k) calculated across coders are reported; for self-report questionnaires, correlations (rs) among family members are listed. Two types of validity data are summarized. Concurrent/predictive validity includes correlations between the measure and outcomes it is purported to predict as well as evidence of the measure’s ability to differentiate between distinct groups. Convergent validity includes correlations between the reviewed measure and other measures assessing the same or similar constructs.

**General Family Functioning**

Thirteen measures of general family functioning were reviewed. Many of the measures could be traced back to specific models of family functioning. Measures originating from the same model are clustered together in Supplementary Table S2, forming seven groups of instruments (some only include a single measure). Table S2 presents an overview of the theory on which the measures are based, their format, the specific constructs assessed (i.e., subscales and specific areas coded), and other information (e.g., applicable age ranges and additional versions).

Four well-specified models of family functioning (The Beaver’s Systems Model, The Circumplex Model of Marital and Family Systems, The McMaster Model of Family Functioning, and The Process Model of Family Functioning) underlie eight of the reviewed measures. Observational coding systems and self-report questionnaires have been developed for each model. A Clinical Rating Scale is associated with The Process Model of Family Functioning, but it has not been presented in any empirical peer-reviewed paper, so was not reviewed. The remaining measures include two observational coding systems (The Iowa Family Interaction Rating Scales, IFIRS; and The System for Coding Interactions and Family Functioning, SCIFF) and three versions of one widely used self-report questionnaire (The Family Environment Scale, FES). The three versions of the FES reviewed include: (a) the original scale; (b) the Family Relationship Index (the first three subscales); and (c) the Kronenberger and Thompson (1990) factor-analytic component scoring. These additional versions of the FES are used in pediatric populations and have separate empirical bases, so were reviewed separately from the original FES. In total, six observational coding systems and seven questionnaires were reviewed.

**Observational/Clinician-rated Interviews**

Five of the six observational coding systems reviewed qualified for a “well-established” rating. They include: The Beavers Interactional Scales (BIS), The McMasters Clinical Rating Scale (MCRS), the Mealtime Interaction Coding System (MICS), The Circumplex Clinician Rating Scale (CCRS), and The Iowa Family Interaction Rating Scales (IFIRS). The System for Coding Interactions and Family Functioning (SCIFF) qualified for an “approaching well-established” rating.

Information regarding the BIS (Beavers & Hampson, 2003) is primarily reported in nonpeer-reviewed venues (i.e., manuals, books, and chapters) or nonempirical peer-reviewed manuscripts (i.e., reviews and commentaries). These measures (the Competence and Style scales) do, however, meet the guidelines for “well-established” assessment devices. Various groups have published using the measures, the manual is available, and the psychometric properties reported are typically adequate. Although not widely used in pediatric populations, these scales (or versions of them) have been used to assess family functioning in families with children with anorexia nervosa (Wallin, Roijen, & Hansson, 1996), infants with colic (Raiha, Lehtonen, Huhtala, Saleva, & Korvenranta, 2002), and children with diabetes (Bobrow, AvRuskin, & Siller, 1985). Raiha and colleagues reported excellent
inter-rater reliability in their study of families with infants with colic (κ = .79–.83); the other authors did not report psychometrics.

The McMaster Model of Family Functioning has produced two measures relevant to this category and both qualify for “well-established” ratings. The MCRS (Miller, Kabacoff, Epstein, & Bishop, 1994) is a coding system designed to be used with the McMasters Structured Interview of Family Functioning (McSiff) to assess problem solving, communication, roles, affective responsiveness, affective involvement, behavioral control, and general functioning of families. The MCRS has been used in families of children with traumatic brain injuries with evidence of high inter-rater reliability (ICCs $= .76–.96$) and moderate correlations with the Family Assessment Device ($r = .38–.62$; Barney & Max, 2005; Max et al., 1997; Max, Lindgren, Pearson, Ihrig, Welborn, 1998).

The MICS was not developed by the McMaster’s group, but is strongly based upon the model (Dickstein, Hayden, Schiller, Seifer, & San Antonio, 1994). The MICS involves rating a videotape of a meal at the family’s home. It has been used repeatedly in investigations of families of children with cystic fibrosis with acceptable to high inter-rater reliability (ICCs $= .64–.93$) across subscales and studies (Janicke, Mitchell, & Stark, 2005; Mitchell, Powers, Byars, Dickstein, & Stark, 2004; Spieth et al., 2001).

The CCRS, based upon the Circumplex Model of Marital and Family Systems (Olson & Gorall, 2003) also qualifies for a “well-established” rating. Families are coded on adaptability, cohesion, and communication. Information about the CCRS is primarily available through nonpeer-reviewed reports (e.g., the manual, chapters, and commentaries); however, at least two groups of investigators have published empirical peer-reviewed papers reporting adequate psychometric properties of this observational coding system. We found no evidence of the CCRS being used in pediatric samples.

The IFIRS (Melby & Conger, 2001) is a widely used observational measure in family and developmental psychology, applicable across ethnic and racial groups for families of children aged 2 years and older. Composite scales are created from individual subscales (listed in Table S2) and Table S1 presents a range of some of the more frequently developed composite constructs. Though reported ICCs have sometimes been low (i.e., $< .50$), the majority of the information regarding the psychometric properties of this system qualifies it for a “well-established” rating. We uncovered one report using the Dyadic Relationships and Group Problem Solving scales of the IFIRS with families of children with cystic fibrosis (DeLambo, Levers-Landis, Drotar, & Quitter, 2004). Inter-rater reliabilities were adequate (Spearman’s $r = .76–.86$ across scales), family relationship quality scores were highly correlated across dyads ($r = .80–.93$) and CF-specific and CF-nonspecific problem solving were moderately associated ($r = .56$).

The SCIFF was rated as “approaching well-established.” This observational scale has been used by a few groups publishing in peer-reviewed venues, and the manual is available; however, the only complete account of its psychometric properties has been presented in a nonpeer reviewed chapter (Lindahl & Malik, 2001). Although a few respondents to the Div54 listserv survey indicated that they had used or had considered using the SCIFF, publications reporting such data for pediatric samples were not found in the literature.

**Self-report Questionnaires**

Of the seven self-report measures of general family functioning reviewed, three qualified for “well-established” ratings. These included the Family Assessment Device (FAD), the Family Assessment Measure-III (FAM-III), and the Family Relationship Index (FRI) of the Family Environment Scale (FES). The remaining measures, the Beavers Self-report Family Inventory (SFI), the Family Adaptability and Cohesion Evaluation Scale-IV (FACES-IV), and the FES, original version and Kronenberger & Thompson Factor-analytic Component scoring (FES-K&T) all qualified for “approaching well-established” ratings.

The self-report FAD has a very large evidence base arising from peer-reviewed manuscripts and, from quite early in its development, it has been used in medical populations (Kabacoff, Miller, Bishop, Epstein, & Keitner, 1990). The “well-established” FAD has been used broadly in pediatric samples including families of children with asthma, cancer, cerebral palsy, diabetes, inflammatory bowel disease, sickle cell disease, spina bifida, spinal cord injuries, thalassemia, and traumatic brain injury. Typically, only the 12-item General Functioning subscale is reported, but it has excellent internal consistency reported in the .85–.90 range. When used in pediatric samples, the internal consistency of the other subscales is typically above .70, except for the shorter Roles subscale that has an internal consistency of .60–.69. One report documented an $r$ of .47 for the Problem Solving subscale (Streisand, Kazak, & Tercyak, 2003). Despite its widespread use in pediatric samples, psychometrics properties beyond internal consistency have not been reported in this population.
The “well-established” FAM-III (Skinner, Steinhauer, & Sitarenios, 2000) and its short-form have been used in a variety of pediatric populations including families of children with chronic pain, cystic fibrosis, developmental disabilities, eating disorders, and hemophilia. It has also been used among families with preterm infants. The psychometric properties of this scale specifically within pediatric populations are rarely reported, though internal consistencies for the General Scale of the full version ($\alpha = .95$; Van Riper, 2000) and the short-form ($\alpha = .80$; Trute & Hiebert-Murphy, 2002) have been reported as excellent.

The FRI (Holahan & Moos, 1982) received a rating of “well-established.” This subset of scales from the FES and their composite score has been used with families of children with asthma, diabetes, juvenile rheumatoid arthritis, recurrent abdominal pain, sickle cell disease, and those undergoing bone marrow transplant (BMT). In these samples, coefficient $\alpha$ ranges from .73 to .78 for the total score. FRI scores predict parental distress related to child BMT (Phipps, Dunavant, Lensing, & Rai, 2005), child abuse potential among parents of children with developmental disabilities (Aniol, Mullins, Page, Boyd, & Chaney, 2004), and illness management for adolescents with diabetes (Naar-King, Podolski, Ellis, Frey, & Templin, 2006).

The Beavers SFI (Beavers & Hampson, 1990) qualified for a rating of “approaching well-established.” Various independent research teams have published reports including data from the SFI and the items are readily available. Most of the psychometric properties, however, have been presented in nonpeer reviewed venues (e.g., chapters and commentaries). Additionally, some reports have indicated that the three-item Leadership subscale has very low internal consistencies (e.g., $\alpha$’s = .02–.31), so its use as a separate scale needs to be carefully scrutinized. The SFI has been used in an investigation of families in which parents have epilepsy (Krawetz et al., 2001); however, psychometric properties were not reported.

FACES-IV and its companion scales, Family Communication (FCS) and Family Satisfaction (FSS), are the current self-report questionnaires of the Circumplex model (Olson & Gorall, 2003). Past versions of FACES have not been capable of capturing the curvilinear aspects of the Circumplex model (i.e., cohesion and adaptability were found to be linearly related to adjustment instead of extremes on either end of the continuum predicting maladjustment). FACES-IV includes six subscales: two designed to assess the mid-ranges of adaptability and cohesion, and four new subscales to assess the extremes of these dimensions (rigid, chaotic, disengaged, and enmeshed). Two empirical reports, from two distinct investigatory teams (Craddock, 2001; Franklin, Streeter, & Springer, 2001), provide some psychometric information for a preliminary version of FACES-IV, but complete data have only been presented in nonpeer-reviewed documents. Preliminary psychometric properties are promising, but more information is needed before these measures will qualify for a well-established rating. Previous versions of the FACES instruments have been used repeatedly in pediatric samples; however, data from FACES-IV has not yet appeared in press in these populations.

The FES received an “approaching well-established” rating. The internal consistency estimates of some of the FES subscales have been found to be quite low in some samples. Part of the reason for this is the true–false response format; however, even when a 4- or 5-point Likert scale is used, some subscales have coefficient $\alpha$ that consistently fall below our standard for a “well-established” rating (please refer to Table S1). For a more thorough discussion of these issues, please see Sanford, Bingham and Zucker (1999) or Roosa and Beals (1990) with a reply by Moos (1990). The FES has been used with families of children with arthritis, asthma, burns, cancer, chronic pain, diabetes, epilepsy, end-stage renal disease, hemophilia, physical disability, spina bifida, traumatic brain injury, and Turner syndrome. Within these populations, the following internal consistency figures have been published: Cohesion, $\alpha = .61–.77$; Conflict, $\alpha = .60–.78$; Organization, $\alpha = .55–.60$; Expressiveness, $\alpha = .49–.58$; Control, $\alpha = .42–.58$; and Independence, $\alpha = .44$. Evidence of test–retest reliability has been reported using a sample of mothers of children with severe retardation ($r’s = .38–.79$ across subscales over a 2–3 year time lag; Rousey, Wild, & Blacher, 2002). Also, as in the general population, FES Cohesion, Expressiveness, and Conflict subscale scores relate to behavioral problems among children with cancer (Varni, Katz, Colegrove, & Dolgin, 1996), arthritis (Cuneo & Schiaffino, 2002), and diabetes (Holmes et al., 2006), partially substantiating its validity in these populations.

The Kronenberger and Thompson (1990) Factor-analytic Component scoring of the FES also qualifies for an “approaching well-established” rating. Investigators using this scoring method rarely report the psychometric characteristics of the subscales. In at least one case, data for the Controlling subscale was not analyzed due to low internal consistency ($\alpha = .11$; Brown, Madan-Swain,
& Lambert, 2003). This particular version of the FES was created through factor analysis of data arising from families of children with chronic illness then replicated in a normative sample. It has been used in samples of children with cancer, diabetes, juvenile rheumatoid arthritis, and sickle cell disease.

Summary
Eight of the thirteen reviewed measures of general family functioning qualified for classification as “well-established.” The observational coding scales originating from the Beavers Systems Model (BIS), the three instruments originating from the McMaster Model (FAD, MCRS, and MICS), the observational IFIRS and CCRS, and the self-report FAM-III and FRI were all classified as “well-established.” The self-report SFI, FACES-IV and two versions of the FES (original and Kronenberger & Thompson scoring) were “approaching well-established.”

Dyadic Family Relationships
Eight measures assessing three types of dyadic family relationships were reviewed: (a) marital/partner dyad (three measures); (b) parent–child dyads (four measures); and (c) sibling dyads (one measure). An overview of these measures is provided in Supplementary Table S3.

Marital/Partner Relationships
The measures of marital/partner relationship include the Locke–Wallace Marital Adjustment Test (MAT), The Dyadic Adjustment Scale (DAS), and the Marital Satisfaction Inventory-Revised (MSI-R). All of these measures have a long history in the empirical literature and all were categorized as “well-established.”

The MAT was originally published over 45 years ago with the goal of assessing marital adjustment (Locke & Wallace, 1959). Extensive empirical evidence supports the psychometric properties of this measure, its value in distinguishing distressed and nondistressed couples, and its ability to document predicted changes in marital adjustment (e.g., upon the birth of a first child; outcome after intervention). We uncovered two published reports utilizing this measure within a pediatric psychology context. Both used samples of families with various disorders (Berge, Patterson, & Rueter, 2006; Cahill & Gladden, 1996). The MAT showed evidence of good internal consistency ($\alpha > .77$) and test–retest reliability ($r's > .70$ over 12–15 months).

The 32-item DAS (Spanier, 1976) has four subscales: Dyadic consensus, Dyadic satisfaction, Dyadic cohesion, and Affectional expression. Much debate has been raised about the factor structure of this measure (Kazak, Jarmas, & Snitzer, 1988; Sharpley & Cross, 1982), though more recent analysis indicates that the four subscales contribute hierarchically to a single total score (Sabourin, Lussier, Laplante, & Wright, 1990). Analyses regarding the structure of the DAS led to two popular “short-forms” of the instrument: the Revised Dyadic Adjustment Scale (RDAS), a 14-item scale with three of the original four subscales (Busby, Crane, Larson, & Christensen, 1995) and the Abbreviated Dyadic Adjustment Scale (ADAS or DAS-7), a seven-item single factor measure (Sharpley & Cross, 1982; Hunsley, Best, Lefebvre, & Vito, 2001). Each of these versions of the DAS qualifies for a “well-established” rating. The DAS has been used in studies of families with children with congenital heart disease, epilepsy, spina bifida, and mixed chronic illness conditions. It has also been used in studies of risk for childhood injury. The applicability of the DAS norms in pediatric samples has been an issue of some concern (Walker, Manion, Cloutier, & Johnson, 1992); however, the internal consistency is excellent for the total score ($\alpha > .88$) and it has been found to reliably predict marital distress in these populations (Walker et al., 1992).

Finally, the MSI-R (Snyder, 1997) provides a more detailed assessment of marital functioning. It includes 150 true–false items comprising 10 subscales, a global distress scale, and two validity scales. It was rigorously developed, primarily for clinical applications. The majority of the psychometric data regarding the MSI-R has not been published in peer-reviewed empirical reports and almost all of it comes from the same investigatory team. The measure does, however, fulfill our criteria as “well-established.” We found no evidence of its use in pediatric populations.

Parent–Child Relationships
Three primary self-report measures of parent–child relationships were reviewed: The Revised Children’s Report of Parental Behavior Inventory (CRPBI-30). The Inventory of Parent and Peer Attachment (IPPA), and the Issues Checklist (IC). One observational coding system was also reviewed: The Constraining and Enabling Coding System (CECS). Each of these measures qualified for a rating of “well-established.”

The CRPBI-30 (Schludermann & Schludermann, 1988) was designed to assess children’s perspectives of their parents’ parenting behavior through the administration of 30 items. It is the latest iteration of a 260-item scale first published in 1965 (Schaefer, 1965) and is derived from a 108-item version (Schludermann & Schludermann, 1970). The 30 items fall onto three subscales: Acceptance, Psychological Control, and Firm
Control/Discipline. Overall, the psychometric properties of this measure are acceptable resulting in a “well-established” rating, even though its development has not been presented in a peer-reviewed publication. The CRPBI-30 has not been used in any published studies involving pediatric populations; however, data from subscales of the 108-item version have been gathered in samples of children with spina bifida with satisfactory internal consistency ($\alpha = .74–.80$; McKernon et al., 2001).

The IPPA was designed to assess perceived attachment with primary figures in adolescents’ lives. Soon after publication, the authors refined the measure to a 75-item version assessing attachment to mothers (25 items), fathers (25 items), and peers (25 items) through use of three subscales: Trust, Communication, and Alienation (Armsden, & Greenberg, 1987). The IPPA (Mother, Father, and Peer Version) has been used in families with adolescent girls with diabetes. Though reliability figures were not reported, poorer communication and less trust between parents and adolescents was related to adolescent eating disturbances in this population, demonstrating some evidence of validity (Maharaj, Rodin, Olmsted, & Daneman, 1998).

The IC (Robin & Foster, 1989) was designed to measure conflict between parents and adolescents. It is a 44-item list of issues on which adolescents and parents may disagree. Respondents indicate the frequency and anger intensity with which each topic was discussed in the past month. Psychometric properties are acceptable and the scale qualified for a “well-established” rating; however, there is often significant discrepancy between adolescent and parent reports. It has been used with adolescents with type I diabetes and those at risk for HIV exposure. Shorter and specific versions have been developed, such as a 15-item version for families of children with cystic fibrosis (DeLambo et al., 2004). The original scale demonstrates excellent internal consistency in pediatric samples ($\alpha = .86–.89$; Dashiff, Bartolucci, Wallander, & Abdullah, 2005; Viikinsalo, Crawford, Kimbrel, Long, & Dashiff, 2005).

The CECS (Hauser et al., 1984) is an observational coding system that assesses: (a) constraining interactions, namely those that impede and undermine development; (b) enabling interactions, those that promote growth and development; and (c) discourse change, evidence of the way in which the adolescent responds to the constraining or enabling interaction (e.g., with progression and regression). Data derived from the CECS has been published by at least two separate investigatory teams, the manual is available, and its psychometric properties are sufficient for a “well-established” rating. The CECS has been used to assess families with children diagnosed with diabetes and families of children who are acutely ill. Inter-rater reliability was satisfactory ($\kappa = .42–.93$; 84–99% agreement across items).

Sibling Relationships
The final measure of dyadic relationships, the Sibling Relationship Questionnaire (SRQ; Furman & Buhrmester, 1985), was designed to capture aspects of sibling relationships. It includes 15 three-item subscales falling onto four factor analytically derived scales: Warmth/Closeness (21 items), Relative Status/Power (12 items), Conflict (9 items), and Rivalry (3 items). Many investigatory teams have published reports using the measure as developed or shorter derivations and the psychometric properties of the scale qualify it for a “well-established” rating. The measure has been used in studies of children with cancer (Labay & Walco, 2004) and arthritis (Weiss, Schiaffino & Ilowite, 2001) and their siblings; psychometric properties were not reported.

Summary
Our review of measures of dyadic relationships within families revealed eight “well-established” measures: three self-report measures assessing marital relationships (MAT, DAS, and MSI-R), three self-report (CRPBI-30, IPPA, IC), and one observational measure (CECS) of parent–child relationships, and one measure of sibling relationships (SRQ).

Families in the Context of Childhood Chronic Medical Conditions
The final measures reviewed were developed specifically to capture aspects of family functioning within pediatric psychology contexts. Based upon the constructs these measures were developed to assess, they can be divided into three categories: (a) normative family processes (e.g., separation and individuation) in families of children with chronic health conditions; (b) the impact of childhood chronic health conditions on the family; and (c) family/parental coping with childhood illness or disability. An overview of these measures is provided in Supplementary Table S4.

Normative Family Processes
The Autonomy and Intimacy Rating System (AIRS; Maharaj, Rodin, Connolly, Olmsted, & Daneman, 2001) is a macroanalytic coding system based upon the CECS described earlier. It is designed to measure: (a) intimacy;
(b) adolescent expression of autonomy; and (c) mother facilitation of autonomy. These three subscales have been confirmed by factor analysis and the psychometric properties of the scales are exceptional. To date, only one investigative team has published using the measure in mother–daughter dyads where the daughter has diabetes, qualifying it for an “approaching well-established” rating.

Impact of Childhood Chronic Health Conditions on the Family

Three scales were reviewed that assess the impact of the child’s chronic health condition on the family or parents. These were The Impact on Family Scale (IOF), the Parents of Children with Disability Inventory (PCDI), and the Pediatric Inventory for Parents (PIP). All were deemed “well-established.”

The IOF is one of the earliest developed measures to assess the impact of a child’s chronic illness on the family (Stein & Riessman, 1980). The original form consisted of 27 items plus six sibling-related items. Twenty-four of the items loaded on four factors: Personal Strain/Distress, Disruption of Familial/Social Relations, Financial Burden, and Mastery/Coping. It has been used with parents of children with asthma, cerebral palsy, chronic pain, congenital anomalies, HIV infection, juvenile rheumatoid arthritis, liver failure, and traumatic injury. More recent analysis suggests that across diverse samples, only one emerging factor is consistent—a 15-item Negative Impact on the Family scale (Stein & Jessop, 2003). The authors recommend using this shortened version of the scale that is highly correlated with the Total Impact Score from the previous version (r’s > .97). Data using the new criteria have been published by at least two investigatory teams and the psychometric properties are excellent, qualifying the revised measure, in addition to the original version, for a “well-established” rating.

The PCDI (Noojin & Wallander, 1996) assesses disability-related stress of parents of children with disabilities. This 40-item scale includes four domains confirmed by factor analysis: Medical and Legal Concerns, Concerns for the Child, Concerns for the Family, and Concerns for the Self. The measure has acceptable psychometric properties and qualifies for a “well-established” rating. It has been used with families of children with spina bifida, cerebral palsy, diabetes, and obstetrical brachial plexus injuries.

The PIP (Streisand, Braniecki, Tercyak, & Kazak, 2001) assesses parenting stress specific to caring for a child with an illness. Four domains are assessed: Communication with Medical Team, Emotional Distress, Medical Care, and Role Function. Originally developed with families of children with cancer, the measure has been used in families of children with sickle cell disease, type 1 diabetes, and short stature with no needed modifications and excellent psychometric properties. It is a “well-established” measure.

Family/Parental Coping

The way in which the family copes with childhood chronic conditions is the focus of the final measures in this review: The Coping Health Inventory for Parents (CHIP), the Family Coping Coding System (FCCS), and The Family Narrative Consortium Coding Scheme (FNC).

The CHIP (McCubbin et al., 1983) is a parent self-report inventory including 45 specific behaviors in which one might engage to cope with having a child with a serious illness. Through factor analysis, three coping patterns were uncovered: (a) Family Integration, Cooperation and an Optimistic Definition of the Situation; (b) Maintaining Social Support, Self-esteem, and Psychological Stability; and (c) Understanding the Health Care Situation through Communication with Other Parents and Consultation with the Health Care Team. Some investigatory teams have reported lower internal consistencies for the subscales than the instrument’s authors, but the psychometric properties are adequate overall, qualifying this measure for a “well-established” rating. It has been used with parents of children with cancer, cerebral palsy, cystic fibrosis, diabetes, end-stage renal disease, epilepsy, muscular dystrophy, and spina bifida.

The FCCS (Hauser et al., 1984) is an observational coding system designed to describe and differentiate the ways that families cope with a stressful life event. All of the studies using this methodology, however, have examined the stressor of having a child with a medical condition (i.e., diabetes, juvenile rheumatoid arthritis, and acute illness/injury). The coding system focuses on the coping strategies of the family as a unit, as opposed to the coping strategies of individuals. This system has been used and data published by two distinct investigatory teams, and the manual is available; however, current reports of the psychometrics are vague and modest at best qualifying this measure for an “approaching well-established” rating.

The FNC Coding Scheme (Fiese et al., 1999) is a coding system of the narratives of families. It is designed to provide insight into how the family derives meaning from stressful events and incorporates them into their
worldview. The FNC Coding Scheme has been used in families of children with asthma, adopted adolescents, and depressed mothers. The reported psychometric properties are excellent, and data derived from the coding scheme have been presented in multiple peer-reviewed papers, however, only by one investigatory team, placing the measure in the “approaching well-established” category.

Summary
The measures of families within the context of childhood chronic health conditions chosen for review fell into three broad categories: (a) assessment of normative family processes (one observational measure); (b) assessment of the impact of childhood chronic conditions on the family (three self-report indices); and (c) assessment of family/parent coping (one self-report and two observational measures). All of the self-report measures (IOF, PCDI, PIP, and CHIP) qualified for “well-established” ratings; the three observational coding systems (AIRS, FCCS, and FNC) qualified for “approaching well-established” ratings.

Discussion
This review selected 29 family measures deemed relevant to the field of pediatric psychology and after careful review of their empirical bases, determined that 19 (66%) were “well-established.” The remaining ten instruments were deemed “approaching well-established.” These findings are encouraging for the field of pediatric psychology, providing many excellent options when choosing a family assessment measure.

Strengths of the Measures Reviewed
The measures reviewed have many positive qualities. The “well-established” measures are supported by peer-reviewed psychometric data. Many of the reviewed instruments derive strength from their basis in theoretical models. Furthermore, some of the best developed models have generated a combination of observational coding systems, clinician-rated interviews, and self-report measures allowing for multiple perspectives when assessing the family. Many of the measures reviewed have been adapted for use with linguistic minorities and peoples of other cultures (e.g., FAD, IFIRS, SCIFF, MAT, DAS, MSIR, CRPBI, IOF, and CHIP). A subset of the family measures has been developed specifically for use in pediatric populations and these self-report instruments demonstrate strong psychometric properties across a range of pediatric samples. Observational measures for use in pediatric populations are gathering additional empirical data and may soon achieve “well-established” ratings.

Limitations of the Measures Reviewed
This review determined that some widely-used family measures have psychometric properties that are less than ideal or simply not well documented in pediatric samples. This is a serious limitation that may lead to instability of findings or even erroneous conclusions in our work. There are also practical limitations to using some family measures within pediatric populations. Some measures are long (e.g., 150 items), intrusive (e.g., videotaping a family meal), and/or time intensive (e.g., transcribing audiotapes and coding speech acts), which may be particularly problematic when trying to investigate or intervene with regard to a stressful event such as a newly diagnosed illness or an intensive medical regimen. Also, observational coding systems may be subject to a specific limitation when applied in pediatric samples—the coder may be able to detect the child’s medical condition and this might lead to bias in their ratings. It is important to understand how these issues may influence family assessment in our field. Also, there is great variability across the types of health conditions with which we are concerned and therefore, measures may be limited in their applicability within the field, even if they were developed for use with children with chronic health conditions.

Limitations of this Review
Reviewing measures for a field as broad as family assessment is a daunting task and there are limitations to our review. First, we needed to limit the scope of the review significantly. Many types of family-related measures relevant to pediatric contexts (e.g., measures of attachment and general parenting stress) were not included to ensure adequate coverage of measures within the domains we selected. Additionally, we do not claim to have assessed every instrument within the selected areas. We used multiple methods of identification (i.e., listserv survey, review of the literature, and collective expertise) to decrease this possibility, but potentially relevant and useful measures were likely to be overlooked if used infrequently. Additionally, it is difficult to ascertain the validity of the information collected through the listserv survey.
Like all review endeavors, ours is also limited by reliance upon the available evidence base. If a measure has not performed reliably and validity within a pediatric context that information may not be published. This may lead to a positive bias in our review.

Finally, in applying the SPP Task Force criteria, we found it necessary to further define the terminology. We attempted to be very clear in our interpretation and application of the standards. Still, we found that there was much variability in the extent and quality of the evidence base for measures of a given classification level. More precision could be gained and more information provided to our readers, if the system included more levels of discrimination (especially at the upper end of the continuum) or if measures could receive separate ratings for different aspects (e.g., internal consistency and generalizability) of their evidence base.

**Recommendations**

Family assessment is a challenging endeavor, yet vital in the pursuit of increasing our knowledge within pediatric psychology, both clinically and in research. To advance the evidence base of family measures relevant to the field of pediatric psychology, we recommend the following. Primarily, more attention must be paid to the psychometric properties of family measures when used in pediatric populations. Most of the family measures used in pediatric psychology were developed within the general population and they have been applied in pediatric samples without investigation of their reliability and validity within these specific samples. As a minimum requirement, researchers should report the internal consistency of the responses collected on the measure for the sample on which they are reporting. As demonstrated in this review, regarding the FES, internal consistency is not a stable characteristic of a measure and may vary from one sample to the next. Investigating and reporting the Cronbach $\alpha$-coefficient alerts others to potential problems with data from the measure in specific populations.

In addition to internal consistency, other forms of reliability should be investigated and shared. For example, test–retest reliability of family measures was rarely reported for pediatric samples, even when the data were available. Also, correlations between reports of different respondents were rarely reported. Such data are not only valuable information about different perceptions of the same family’s functioning, but they might illustrate interesting ways in which families of children with chronic health conditions differ from the general population.

Investigations of the validity of family measures within pediatric samples and information on norms for these samples are also needed. Families of children with chronic health conditions may score in “unhealthy” or “dysfunctional” ranges using norms and cut-offs developed in the general population; however, these patterns of functioning may actually be adaptive within these families. For example, rigidity with regard to roles and rules may be seen as problematic in the general population; yet, for families needing to adhere to a complex treatment regimen for their child, greater rigidity may be associated with better medical outcomes. Similarly, families of children with chronic health conditions may be able to endure different levels or patterns of distress, while still meeting the developmental needs of their members. These types of questions of the validity of family measures are rarely investigated.

Lastly, on this point, it should be noted that when the measures we reviewed received a rating of “approaching well-established,” it was often because the psychometric properties of the instruments had not been reported in peer-reviewed empirical outlets. Developers of family measures need to attend closely to the psychometric properties of their measures, submit their work for evaluation by their peers, and strive to publish in empirical venues.

Next, we have specific recommendations regarding methodology, which may help inform family assessment. First, attention needs to be paid to all members of the family. Family members may perceive the same event differently and be affected in distinct ways. Despite repeated calls within our field, fathers are still commonly not involved in pediatric psychology research and treatment (Phares, Lopez, Fields, Kamboukos & Duhig, 2005) and siblings rarely receive any attention. Recent recommendations have been made regarding ways to integrate data from multiple family members in either research or treatment contexts (Atkins, 2005; Cook & Kenny, 2004; Dekovic & Buist, 2005). Assessing all family members will provide more data on the psychometric properties of the measures and help refine and improve our methods.

Second, attention needs to be paid to the heterogeneity of families. Families differ in terms of their structure, member composition, and developmental stage. Shifting demographics suggest that more children are being raised in households with single parents or in blended families with stepparents, stepchildren and
half-siblings, or foster families (Teachman, Tedrow, & Crowder, 2000). In some families, “extended” family members, fictive kin, or friends may actually be part of the primary family unit. Additionally, young or new families are quite different from more established families and may respond quite differently to the challenge of a pediatric condition. Our family assessment methods are challenged by these issues, but valuable suggestions regarding ways to ensure assessment quality given these circumstances have recently been published (Hofferth & Casper, 2007). Further, the ethnic, cultural, and linguistic backgrounds of families are becoming more diverse in America and children at greatest psychosocial and physical risk are more likely to be raised in poverty by parents with limited educational backgrounds (Repetti, Taylor, & Seeman, 2002). Variation in the backgrounds of families may result in lower reliability and validity of measures developed using samples of English-speaking, middle-class, white, and two-parent families. Issues of measure equivalence across diverse sociocultural groups are important and may be advanced through the framework of item response theory as discussed by Bingenheimer, Raudenbush, Leventhal and Brooks-Gunn (2005).

Our third area of recommendation is a call for more attention to the clinical usage and value of family assessments in the published literature. Clinicians and clinical researchers need to know if family measures can be used to guide clinical interventions, if they are predictive of treatment outcomes, and if they are sensitive to change. Among the measures reviewed, the FAM-III, FES, SFI, MAT, MSR-I, IC, and CHIP are all documented as being valuable in predicting or assessing the outcomes of treatment. Other measures may also be valuable in this regard, but such evidence was not uncovered in the available literature.

Conclusions

When faced with the challenge of choosing a family measure, clinicians and researchers have many questions to ask themselves. They must consider which domain of functioning to assess, which components of the family system to assess, what level of specificity to choose, and which perspectives/methods will be most useful to address their specific questions. The current review provides information regarding the evidence base of measures of family functioning and dyadic relationships and reveals that despite some limitations, the majority of the measures described in this review are valuable sources of information and merit continued exploration in pediatric psychology contexts.

Supplementary Data

Supplementary data are available at JPEPSY Online.

Acknowledgments

Preparation of this article was supported in part by a grant to the first author from the American Cancer Society (MRSG 0521301). The authors express their gratitude to Lauren Bradley for her extensive library work and Rowena Conroy, PhD, Lynne Kaplan, PhD and Tracy Hills, BA for their contributions. We also thank Lindsey Cohen (Chair) and the APA Division 54 Assessment Task Force Committee for guidance during this project.

Conflict of interest: None declared.

Received November 11, 2006; revisions received August 4, 2007; accepted August 21, 2007

References


McKernon, W., Holmbeck, G., Colder, C. R., Hommeyer, J. S., Shapera, W., & Westhoven, V. (2001). Longitudinal study of observed and perceived family influences on problem-focused coping


