

Children's family experiences: development of the PROMIS[®] pediatric family relationships measures

Katherine B. Bevans¹ · Anne W. Riley² · Jeanne M. Landgraf³ · Adam C. Carle⁴ · Rachel E. Teneralli⁵ · Barbara H. Fiese⁶ · Lisa J. Meltzer⁷ · Anna K. Ettinger^{8,9} · Brandon D. Becker⁵ · Christopher B. Forrest^{5,10}

Accepted: 15 June 2017 / Published online: 22 June 2017
© Springer International Publishing AG 2017

Abstract

Purpose To describe the development of pediatric family relationships measures, with versions for child self-report (8–17 years) and parent-report for children 5–17 years old. Measures were created for integration into the Patient Reported Outcomes Measurement Information System (PROMIS[®]). **Methods** Semi-structured interviews with 10 experts, 24 children, and 8 parents were conducted to elicit and clarify essential elements of family relationships. A systematic literature review was conducted to identify item concepts representative of each element. The concepts were

transformed into items that were iteratively revised based on cognitive interviews ($n = 43$ children) and item translatability review. Psychometric studies involving 2846 children and 2262 parents were conducted to further refine and validate the instruments.

Results Qualitative procedures supported the development of content valid Family Relationships item banks. Final child- and parent-report item banks each contain 47 items. Unidimensional item banks were calibrated using IRT-modeling to estimate item parameters representative of the US population and to enable computerized adaptive test administration. Four- and eight-item short forms were constructed for standard fixed format administration. All instruments have strong internal consistency, retest-reliability, and provide precise estimates of various levels of family relationship quality. Preliminary evidence of the instruments' validity was provided by known-group comparisons and convergence with legacy measures.

Conclusion The PROMIS pediatric Family Relationships measures can be applied in research focused on determinants, outcomes, and the protective effects of children's subjective family relationship experiences.

Electronic supplementary material The online version of this article (doi:10.1007/s11136-017-1629-y) contains supplementary material, which is available to authorized users.

✉ Katherine B. Bevans
katherine.bevans@temple.edu

¹ Temple University College of Public Health, 1700 N Broad St, 301-H, Philadelphia, PA 19121, USA

² Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

³ HealthAct, Boston, MA, USA

⁴ Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

⁵ The Children's Hospital of Philadelphia, Philadelphia, PA, USA

⁶ University of Illinois at Urbana-Champaign, Champaign, IL, USA

⁷ National Jewish Health, Denver, CO, USA

⁸ Booz, Allen, Hamilton, Mclean, VA, USA

⁹ Johns Hopkins University, Baltimore, MD, USA

¹⁰ University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

Keywords Child · Family relationships · Family belonging · Family interactions · Social health · Child-report · Parent-report

Children's family relationships powerfully shape their current and future health and quality of life [1–3]. More precise measures are needed to better characterize the types of family experiences that promote and maintain health across the life course. Existing measures focus primarily on how families function. They are less useful for characterizing the complexity of family social interactions,

especially during adolescence when youth become increasingly independent and differentiated from their families. Moreover, most measures assess family relationships from parents' perspectives. Even those that include youth-report versions (e.g., FACES-IV [4]; family environment scale (FES) [5]) ask respondents to describe the perspectives and feelings of 'the family,' rather than their own viewpoint. Many questionnaires lack a specific recall period despite evidence that family relationships, interactions, and functioning change over time. As health research increasingly addresses the family context, improved measures are needed to support understanding of normative developmental change in family relationships and the impact that families have on children's quality of life.

The Patient Reported Outcome Measurement Information System (PROMIS[®]) is a federal initiative that aims to develop and standardize person-reported outcome (PRO) measurement tools for clinical research and application [6, 7]. PROMIS tools are developed in accordance with rigorous, mixed-method scientific standards [8–10]. Methods include item response theory (IRT) modeling to inform the selection of items that reliably assess a health outcome with minimal gaps and item redundancies, and enable detection of measurement bias via differential item functioning (DIF) and development of computerized adaptive test (CAT) algorithms [11–13]. PROMIS pediatric instruments are developed for children aged 8–17 and parents of children aged 5–17.

This paper describes the development and psychometric evaluation of the PROMIS[®] pediatric family relationships instruments. In the PROMIS model of health, family relationships are elements of Social Health, which "encompasses the ways in which individuals connect with important others, including communication, companionship and understanding, and the quality, reciprocity and size of an individual's social network" [14].

Methods

The family relationships measures were initially developed based on literature reviews; concept elicitation with experts, parents, and children; cognitive interviews; translatability assessments; and reading level analysis. Item banks and short forms were refined based on psychometric analyses of data collected in two large field studies involving diverse samples of children and parents. Figure 1 summarizes the instrument development process.

Conceptualization

Initial conceptualization of family relationships was informed by theories of attachment [15, 16], bio-ecological

influences on health [17], and living systems [18]. Attachment theory posits that children's beliefs about family members' caring, predictability, and dependability exert profound and lasting effects on development, health, disease, and mortality [15, 16]. The security of children's family connections impacts psychobiological processes of brain development and gene expression; emotional and behavioral regulation; interpersonal competence; and vulnerability to disorders [19–24]. Bioecological theory posits that *proximal processes*, children's frequent, reciprocal interactions in their most immediate environments, powerfully shape their development. As described in living systems theory of health [18] and the family systems literature [25], continuous dynamic person-environment interactions influence neurodevelopment and stress regulation systems so that the environment 'gets under the skin' to foster cumulative advantage or disadvantage in health over time [26]. According to a family systems framework, well-functioning families offer stability and coherence, providing members with a secure base that fosters exploration of the environment and independence [27].

Expert concept specification interviews

The Family Relationships conceptual framework was iteratively refined through semi-structured interviews conducted with ten family and clinical researchers, each with a minimum of three peer-reviewed publications on family relationships. Experts provided feedback on an initial Family Relationships definition and proposed component concepts (see Online Resources 1 for expert interview guide).

Child and parent concept elicitation interviews

Child and parent concept elicitation interviews were conducted to further refine the Family Relationships conceptual framework. Children aged 8–17 ($n = 24$) and parents of children aged 5–17 ($n = 8$) were recruited from urban clinics and community settings serving low/middle income families and higher-income suburban clinics. Children with an intellectual or developmental disability that inhibited them from completing an interview were excluded. Interviewees were asked to describe their feelings about being part of a family, typical family activities, and the impact of family relationships on their/their child's quality of life (see Online Resources 2 for child/parent interview guides). Interview topics were explored until saturation was achieved. Saturation was achieved when a family relationships concept was elicited in at least one but not the last interview and if enough information was elicited to fully understand the meaning and importance of the concept to children [28].

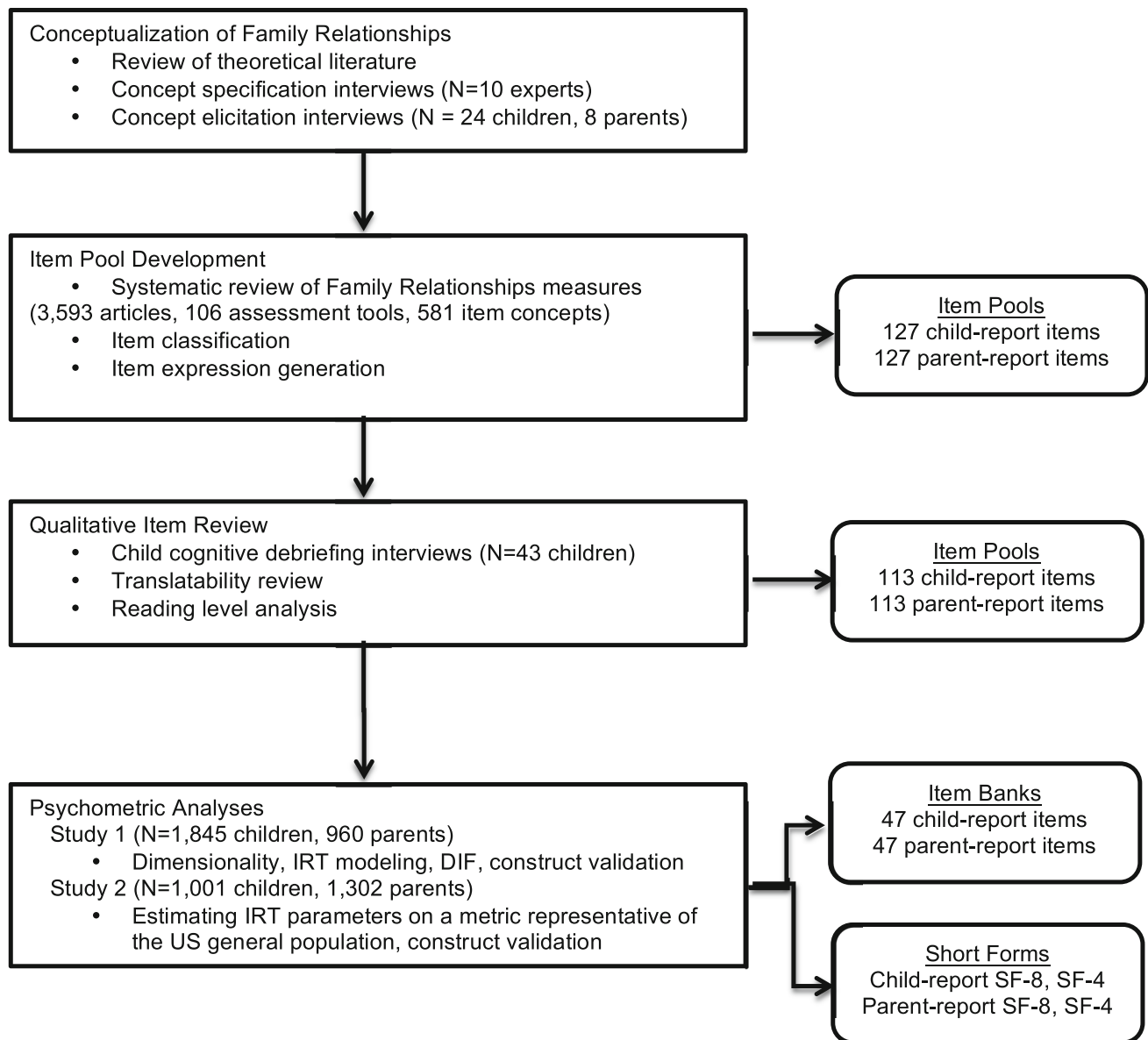


Fig. 1 PROMIS® pediatric family relationships instrument development process

Systematic review of family relationship measures

A comprehensive systematic literature search [29] was conducted using MEDLINE, CINHALL, PsychINFO, and HaPI databases to identify child- and parent-report measures of family relationships. Search terms were constructed to capture conceptual framework components and child- and parent-report measures of family relationships. Search terms included concepts identified in expert, child, and parent interviews (See Online Resources 3). Titles and abstracts were reviewed; tools were obtained from instrument developers, and a catalogue of the concepts they assessed was created.

Classification of item concepts and item expression generation

Three investigators reviewed all item concepts related to family relationships and used a binning and winnowing process to sort them into the eight facet categories and to eliminate redundant or vague concepts [8]. Each item concept was then transformed into child- and parent-report item expressions (questionnaire items) by generating context, stem, and response options consistent with PROMIS standards [10]. Child-report items were written to assess children's perspectives on their families (e.g., "I felt I had a strong relationship with my family.") Most parent-report

items are written to assess parents' perspectives on their child's sense of family relationship quality (e.g., "My child felt he/she had a strong relationship with our family"), but a few assess parents' sense of relationships among family members (e.g., "Our family and my child had fun together"). Complete child- and parent-report items are presented in Online Resources 4–5.

Item understandability

Children aged 8–17 ($n = 43$) participated in cognitive interviews to evaluate and improve item comprehensibility [30]. Children were recruited from urban and suburban clinics and community settings serving a range of families. Children with a disability that inhibited them from completing a cognitive interview were excluded. Each of 127 items was tested with at least five children. Interviewers rated children's comprehension for each item on a scale of 1 (low) to 3 (high) and noted reasons for misunderstandings. Items with average ratings <2 were removed or revised and retested using the same procedures.

Translatability review

Translation experts reviewed each item to identify idiomatic expressions, complex sentences, and concepts that are not easily translated into Spanish and German. Items were translated following a universal language approach following recommendations of the ISPOR PRO Outcomes Translation and Linguistic Validation Task Force [31, 32] and the International Quality of Life Assessment Project [33]. A forward–backward–forward translation approach was applied coupled with family relationship expertise to achieve optimal conceptual and cultural equivalence between the English and other-language versions [31].

Item bank development and psychometric analyses

Two large-scale field tests were conducted to further refine the English-language versions of the instruments using psychometric analyses. In Study 1, child and parent responses to 113 items were subjected to psychometric analyses and IRT-based calibration to finalize the Family Relationships item bank and short forms. In Study 2, the IRT-based item parameters were adjusted to more fully represent the US general population. Both studies included evaluations of the instruments' construct validity.

Participants

Study 1 participants were 1845 children aged 8–17 and 960 parents of children aged 5–17. Just over half (children: 56%, parents: 51%) were recruited from an opt-in online

panel (Op4G). Other participants were recruited through children's schools (children: 38%, parents: 40%) or primary or specialty care clinics (children: 6%, parents: 9%). A subset of child–parent dyads ($n = 23$) from the opt-in panel were randomly selected to complete the items twice, 2 weeks apart for test–retest reliability. The Study 2 participants (1001 children, 1302 parents) were recruited from an online panel of a representative random sample of the US population (GfK Knowledge Panel). For both studies, participants were excluded if they had an intellectual or developmental disability that prevented them from responding to questionnaire items.

Measures

Demographics and chronic conditions Children's and parents' reports of their own/their child's gender, age, race, and ethnicity were used in the analysis of child-report and parent-report measures, respectively. Parents indicated whether their child had a chronic condition that 'lasted or was expected to last at least 12 months, AND interferes with his/her activities.'

PROMIS pediatric family relationships measure Study 1 participants answered 113 family relationships items using 5-point Likert response categories (1–Never to 5–Always). Negatively worded items were reverse-scored; higher scores indicate better family relationships. Study 2 participants completed 15 family relationships items that were identified as candidates for short form versions of the measure in Study 1.

Legacy measures Study 1 child participants completed the NIH Toolbox Emotional Support scale, Maternal Relationships scale, and Paternal Relationships scale [34]. Study 2 parent participants completed the FACES IV balanced cohesion and flexibility subscales [4].

Procedures

Study 1 For the Internet panel, adult participants in an online research panel (Op4G) known to have children aged 5–17 were notified by email of their eligibility. Parents who consented were emailed a link to the online parent questionnaire. After completion, parents of children aged 8–17 were instructed to ask their child to complete the child-report measures. Participants in the test–retest subsample were re-contacted and re-assessed 2 weeks later.

For the school sample, data were collected in three school districts in New Hampshire, Vermont, and Texas selected to diversify the sample with respect to race, ethnicity, socio-economic status, and geographic location. English-speaking students in grades 3–12 with parental

consent were administered questionnaires at school in groups of 25–150 students. Parent questionnaires were sent home with students with postage paid envelopes; parents returned completed surveys directly to the investigators.

For the clinic sample, parents of children attending primary and specialty care clinics and the emergency department at the Children's Hospital of Philadelphia were approached. If interested, they and their child completed the items during their medical visit or at home using the Internet-based NIH Assessment Center data collection platform.

Study 2 The study 2 sample was recruited from GfK Knowledge panel, an existing dual frame (random-digit dial and address-based) online probability panel [35, 36]. Parents with informed consent were emailed a link to the online parent questionnaire. After completing their measures, parents asked their child to complete the child-report questionnaires. Data collection continued until age and gender quotas were met; however, population representativeness for other socio-demographic characteristics was not achieved through quota sampling. Therefore, weights were used to make the online panel representative of the US non-institutionalized population. Initial weights were iteratively adjusted (raked) until the weighted sample's distributions of gender, age, race/ethnicity, education, U.S. Census region, metropolitan area, household internet access, and language (English/Spanish) matched those in the most recent current population survey (CPS) [37].

Psychometric analyses

Using Study 1 data, response frequencies, ceiling and missing percentages, means, and standard deviations were calculated for each item. In accordance with the analytic plan developed for PROMIS instruments, we sought to generate a unidimensional measure of family relationships [38]. Item bank unidimensionality was evaluated using exploratory (EFA) and confirmatory (CFA) factor analyses using Mplus 7.2 and a weighted least squares means and variance adjusted estimator. The EFAs used geomin rotation. Unidimensionality was supported in EFA if a single factor explained a large share of the variance and in CFA by adequate single-factor model fit based on the Comparative Fit Index (CFI), Tucker–Lewis fit index (TLI), and root mean square error of approximation (RMSEA) [39]. Items with CFA factor loadings <0.70 were considered for removal. Items were considered locally dependent if constraining the residual correlation between the items to zero resulted in poor fit, as identified through modification indices [40]. When observed, one of the two locally dependent items was removed. Graphs of item mean scores conditional on the total test score minus the item score

were examined to confirm item monotonicity (i.e., the probability that item endorsement increases as the measured trait increases). Non-monotonic items were removed.

We also probed for DIF. DIF refers to the possibility that people with equivalent levels of a latent variable respond differently to items as a function of another variable (e.g., gender). We used Mplus 7.2 and the multiple group CFA method to test for DIF by gender, age (8–12 vs. 13–17), race (White vs. non-White), and ethnicity (Hispanic vs. non-Hispanic) [41]. This method statistically identifies the model across groups and then tests whether constraining the discrimination parameters to cross-group equality results in statistically significantly poorer fit using an omnibus Chi-square test. If the omnibus test suggests DIF, the deterioration in fit due to each item-level constraint is evaluated. Inflated Type I error is controlled using the Benjamini-Hochberg method [42]. A similar process is used to evaluate DIF in the location parameters. DIF's substantive impact is evaluated by examining changes (if any) in the focal group's mean and variance across invariant and partially invariant models, as well as examining differences in individuals' scores estimated across invariant and partially invariant models [41, 43].

Finally, we used a multi-group IRT approach that made use of both study samples to estimate IRT parameters [44]. Analyses were conducted in Mplus 7.2 using maximum likelihood estimation with a logit link. We did this in order to make use of the data from both samples, as well as achieve IRT parameters and scores that could be interpreted relative to the non-institutionalized US 5- to 17-year-old population. The family relationships metric was statistically identified by setting the IRT mean and variance to 0 and 1, respectively, for Study 2 participants (who represent the US population). IRT mean and variance were freely estimated for the Study 1 group. Parameters for items administered in both studies were constrained to equality across study groups. Items not administered in Study 2 were treated as missing for Study 2 participants. For these items, parameters were estimated based on Study 1 data using the same metric as those established for items administered in both studies. Because the family relationships metric was identified using the US representative sample (Study 2), IRT parameters and scores estimated from those parameters can be interpreted relative to the non-institutionalized US population, 5–17 years old. We conducted these analyses using Mplus 7.2 and maximum likelihood estimation with a logit link.

Based on these parameters, items that provided the most discrimination across varying levels of the latent family relationships continuum were selected for short forms comprising eight items (SF-8) and 4 items (SF-4). Although computerized adaptive tests (CATs) using IRT parameters tend to result in the most precise measurement

with the fewest items, [11] static-length short forms are still needed for contexts in which computer-based assessment is not feasible. Short-form unidimensionality was evaluated based on CFA model fit statistics and item residual correlations. Test information functions (TIFs) were plotted to assess each instrument version's precision and reliability across all levels of the family relationships continuum [45]. Internal consistency reliability and test-retest reliability were evaluated using Cronbach's alpha statistic and intra-class correlation coefficients, respectively.

To examine construct validity, we examined the relationships between T -scores and legacy measures. We calculated full bank, SF-8, and SF-4 Bayesian Expected A Posteriori (EAP) theta (Θ) scores in R using the estimated IRT parameters and Firestar (v 1.2.2) [46, 47]. We then transformed these scores into T -scores ($T = \Theta \times 10 + 50$). A score of 50 represents the average level of family relationship experiences for children based on a national sample given our calibration method.

Convergent validity was assessed by calculating Pearson Product Moment correlations between scores on the Family Relationships and legacy measures. Associations between Family Relationships and emotional support, maternal relationships, and paternal relationships were hypothesized to be significant, positive, and moderate-to-strong. For the parent-report legacy measures, correlations were expected to be strongest for Cohesion because it assesses the emotional bonds core to the Family Relationships measures. Flexibility was expected to be significantly, but less strongly related, as it measures the family's role relationships, problem solving and adaptability. Construct validity was further evaluated by assessing known-group differences in family relationships (child age, gender, race, ethnicity, and chronic condition status). Between-group effect sizes ($d > |0.20|$) were considered meaningful [48]. Given normative developmental trends toward greater differentiation from one's family during adolescence and the multiple items about frequency of family activities, family relationships reported by older children were expected to be weaker than those of younger children.

Results

Concept specification interviews with family science experts

Experts found the definition of family relationships to be fairly complete and congruent with their professional experiences, agreeing that children's security and feeling of connection are central to strong family relationships. Experts recommended a few additions to the definition,

most notably an increased emphasis on the concept of "acceptance," the feeling that family members understand and approve of you.

Child and parent concept elicitation interviews

Children of all ages highlighted the importance of love, loyalty, trust, and dependability within a family. For example, an 8-year-old girl noted "Families love and care. They have to always be there no matter what and don't ever give up because it's very important they stick together." Other family relationship experiences seemed to evolve with development. In response to the question "What makes you feel close to your family?" many younger children described parents who give children something or let them do what they desire. Older children were more likely to identify being able to communicate with parents and being "accepted for who I am." Adolescents also articulated that family relationships affect their self worth, behavior, and peer relationships (e.g., how they choose their friends). Most parents agreed with children's characterization of family relationships and added that family connectedness often stems from sharing time and enjoying activities together. Several parents said that clear rules and shared expectations strengthen family relationships.

Family relationships conceptualization and facet definitions

We identified eight facets of family relationships based on expert, child, and parent concept specification interviews: *sense of family* (feeling that one is part of a family), *love and caring* (feeling loved, cared for, and appreciated by family), *value and acceptance* (feeling understood and respected by family), *trust, dependability, and support* (being able to trust and depend on family for help and understanding), *communication* (being able to talk to family members openly and honestly), *enjoyment* (having fun with family), *organization* (arranging family space and activities systematically), and *predictability* (having consistent routines and clear expectations with which family members comply). See Online Resource 6 for final facet definitions.

Systematic literature review

The literature review produced 3593 unique articles that referenced a measure of family relationships. A total of 133 unique child- or parent-report instruments were identified, and permission was obtained from the lead authors to abstract item concepts from 106 instruments (80%).

Table 1 Demographic characteristics of child and parent participants

	Concept elicitation interviews		Cognitive interviews	Study 1 sample ^a		Study 2 sample	
	Child N = 24	Parent N = 8	Child N = 43	Child N = 1845	Parent N = 960	Child N = 1001	Parent N = 1302
Gender, <i>n</i> (%) male	11 (45.8)	3 (37.5)	15 (34.9)	951 (51.5)	495 (51.6)	523 (52.2)	673 (51.7)
Age of child, <i>n</i> (%)							
5–7 years	–	–	–	–	31 (3.2)	–	301 (23.2)
8–11 years	10 (41.7)	–	23 (53.5)	734 (39.8)	327 (34.1)	387 (38.7)	387 (29.7)
12–14 years	7 (29.2)	–	8 (18.6)	645 (35.0)	340 (35.4)	307 (30.7)	307 (23.6)
15–18 years	7 (29.2)	–	12 (27.9)	466 (25.3)	262 (27.3)	307 (30.7)	307 (23.5)
Ethnicity, <i>n</i> (%) Hispanic	0 (0.0)	1 (12.5)	3 (7.0)	270 (14.6)	80 (8.3)	844 (84.3)	230 (17.7)
Race, <i>n</i> (%)							
White	7 (29.2)	1 (12.5)	10 (23.3)		31 (3.2)		301 (23.2)
Black/African American	17 (70.8)	6 (75.0)	24 (55.8)	734 (39.8)	327 (34.1)	387 (38.7)	387 (29.7)
Asian	0 (0.0)	0 (0.0)	6 (14.0)	645 (35.0)	340 (35.4)	307 (30.7)	307 (23.6)
Other	0 (0.0)	0 (0.0)	0 (0.0)	466 (25.3)	262 (27.3)	307 (30.7)	307 (23.5)
Child has chronic condition, <i>n</i> (%) ^b	8 (33.3)	3(37.5)	21(48.8)	146(17.8) ^b	193(20.1)		

^a Test–retest sample included 23 children and 23 parents from the opt-in internet panel

^b parent report variable, percentage of children for whom parent-report data were available ($n = 820$)

Item concept classification, selection, reduction, and generation

Investigators sorted 581 item concepts that were derived from the 106 instruments into the eight Family Relationships facets and eliminated items that were vague or redundant. The resulting domain-facet-item concept map was evaluated for conceptual saturation. Several new item concepts were generated to ensure comprehensive coverage of every facet. A total of 127 item concepts were translated into two item expression versions: one for children and another for parents.

Child cognitive interviews

Of the 127 items subjected to cognitive interviews, 114 (90%) were retained without revision; 3 (2%) were retained after revisions and were cognitively retested, and 10 (9%) were removed because of poor comprehension that could not be mitigated. Examples of poorly understood items are “I felt close to my family because of what we have shared” and “I felt too connected to my family.”

Translatability review

Four items were removed because of poor translatability: “There was a lot of love in my family” and “I felt like my family and I lived in two different worlds” were idiomatic;

“I felt I could not depend on my family” did not fit with a ‘Never’ response; “I did things with my parents,” was removed because it was too vague to translate accurately. Four other items were modified to improve their translatability, a process that also improved the English clarity. The remaining 113 items were tested in the calibration study (Table 1).

Psychometric analyses

Rates of missingness were <2% for all child-report items and <0.7% for all parent-report items. In general, items were positively skewed and as expected, ceiling effects were more common for the parent-report items (see Online Resources 4–5). The single-factor CFA model comprising all 113 child-report items was a poor fit for the data (CFI = 0.90, TLI = 0.89, RMSEA = 0.06) and EFA revealed two prominent factors (eigenvalues = 59.5, 6.5, respectively). Together, these factors accounted for 46.7% of variance in the child-report items. Whereas 85 (factor 1) items described family relationships in positive terms (e.g., “my parents listened to me”), 28 (factor 2) items were negatively worded (e.g., “my family did not want to hear what I had to say”). Subsequent analyses were conducted with factor 1 items only. Twenty-five of the 85 items were eliminated because they had low factor loadings in single-factor CFA models fit to either the child- or parent-report data. Of these, 23 items represented the “organization” and

Table 2 Pediatric PROMIS family relationship item concepts and IRT parameters

Item concept	Child report					Parent report			
	<i>a</i>	<i>a</i> 1	<i>b</i> ₂	<i>b</i> ₃	<i>b</i> ₄	<i>a</i>	<i>b</i> 1	<i>b</i> ₂	<i>b</i> ₃
Strong family relationship (SF4, SF8)	3.62	2.40	-1.73	-0.95	-0.13	4.33	-2.33	-1.42	-0.62
Felt important to family (SF4, SF8)	3.99	-2.39	-1.72	-0.86	-0.08	3.92	-2.49	-1.45	-0.56
Got help I needed from family (SF4, SF8)	3.25	-2.55	-1.89	-0.95	0.01	3.64	-2.33	-1.24	-0.17
Family had fun together (SF4, SF8)	2.80	-2.63	-1.83	-0.72	0.36	2.26	-2.61	-1.16	0.49
Family made me feel good about myself (SF8)	3.40	-2.46	-1.81	-0.83	0.13	3.27	-2.70	-1.46	-0.18
Family treated me fairly (SF8)	2.46	-2.72	-2.02	-0.89	0.23	2.29	-2.37	-1.16	0.26
Parents listened to me (SF8)	2.64	-2.48	-1.81	-0.74	0.30	1.99	-3.51	-1.88	-0.11
Family paid attention to me (SF8)	2.73	-2.48	-1.73	-0.66	0.36	2.48	-2.70	-1.37	0.10
Felt respected by family	2.55	-2.35	-1.72	-0.72	0.29	2.32	-2.74	-1.33	-0.08
Felt important in family	2.48	-2.54	-1.76	-0.95	0.12	4.37	-2.42	-1.43	-0.50
Felt family believed in me	3.42	-2.76	-1.90	-1.05	-0.17	2.82	-2.49	-1.65	-0.63
Family was interested in what I was doing	2.36	-2.52	-1.83	-0.90	0.17	2.85	-2.69	-1.65	-0.55
Felt accepted by family	3.04	-2.52	-1.80	-1.06	-0.20	3.26	-2.56	-1.57	-0.72
Felt understood by family	1.98	-2.69	-1.83	-0.52	0.63	2.58	-2.28	-1.10	0.08
Could depend on family	2.55	-2.54	-2.04	-1.18	-0.27	4.40	-2.37	-1.58	-0.72
Got support I needed from family	3.29	-2.42	-1.91	-1.05	-0.06	3.39	-2.59	-1.45	-0.37
Knew what I could expect from family	1.78	-3.16	-2.35	-1.11	0.21	1.84	-3.23	-1.71	0.02
Could talk someone in family about my problems	2.04	-2.29	-1.69	-0.87	0.05	2.37	-3.05	-1.71	-0.64
Knew family would help me	3.31	-2.55	-2.05	-1.17	-0.38	2.46	-3.21	-2.24	1.17
Could trust people in my family	3.25	-2.54	-1.89	-1.06	-0.07	2.51	-2.79	-1.87	-0.82
Could trust others in family	2.64	-2.52	-1.87	-1.02	-0.03	2.75	-2.34	-1.67	-0.62
Knew family would be there if I needed them	2.90	-2.67	-2.16	-1.28	-0.47	3.53	-2.70	-1.76	-0.95
Felt everyone in family loved me	3.03	-2.42	-1.81	-1.12	-0.38	4.33	-2.35	-1.50	-0.78

“predictability” family relationships facets. One item from each of 8 locally dependent pairs and 5 non-monotonic items were also removed. Single-factor CFA models adequately fit the data for the 47 remaining items (child-report: CFI = 0.95, TLI = 0.95, RMSEA = 0.07; parent-report: CFI = 0.94, TLI = 0.94, RMSEA = 0.07). The rationale for each item’s removal is presented in Online Resource 7.

Discrimination and location parameters for the 47 final items are presented in Table 2. All items contributed to the differentiation of family relationship quality. Item category probability curves (available upon request) showed that for every child-report item, each of the five response categories provided unique information about family relationship quality in an ordered manner, from the lowest (poorest) to the highest (best) category. Location parameters for the child-report items reflect a sizable range of family relationship experiences ($\Theta = -3.20$ to 0.67). For most parent-report items, the lowest two response categories measured the same extremely poor level of family relationship quality. These response categories were collapsed and assigned the same numeric value. Even after recoding, parent-report items measured a wide range of family relationship experiences ($\Theta = -3.23$ to 1.14).

Items that maximized the conceptual coverage and range of precisely measured aspects of family relationships were selected for the 8- and 4-item short forms (Table 2). CFA fit statistics supported short form unidimensionality and all child- and parent-report instruments were found to have good internal consistency and test-retest reliability (Table 3). Test information functions (TIFs) for the full banks and short forms show that the measures are more useful for discriminating among children with poorer family relationships than among those with better family relationships (Fig. 2).

For child report, we observed no statistically significant DIF as a function of gender. With respect to race and age, there were no statistically significant differences in the discrimination parameters but DIF in the location parameters was observed. For ethnicity, we observed statistically significant DIF in one loading and several thresholds. Despite statistically significant differences in some parameters, DIF did not substantively impact scores estimated using the full bank, SF-8, or SF-4. For all grouping variables, scores ignoring and adjusting for DIF were nearly perfectly correlated. Change in a focal group’s mean after adjusting for DIF were less than 0.25 on the *T*-score

Table 3 Descriptive statistics, CFA model fit, reliability, and concurrent validity for the child- and parent-report measures of family relationship

	Child-report			Parent-report		
	Full bank	SF-8	SF-4	Full bank	SF-8	SF-4
Descriptive statistics ^a						
Mean	51.75	51.54	51.74	51.22	51.16	51.41
SD	10.73	9.83	9.48	9.81	8.97	8.59
Minimum	12.86	17.81	20.01	13.13	16.87	20.91
Maximum	70.11	63.91	61.78	69.07	63.12	60.15
Skew	−0.08	−0.41	−0.61	0.05	−0.29	−0.58
Confirmatory factor model fit ^a						
CFI	0.95	0.99	0.99	0.94	0.99	0.99
TLI	0.95	0.99	0.99	0.94	0.99	0.99
RMSEA	0.07	0.11	0.07	0.07	0.07	0.07
Reliability ^a						
Internal consistency (a)	0.99	0.94	0.90	0.98	0.91	0.86
Test–retest reliability (ICC)	0.84	0.82	0.80	0.81	0.78	0.71
Concurrent validation (<i>r</i>)						
Emotional support ^a	0.77	0.74	0.73	0.23	0.23	0.23
Maternal relationships ^a	0.69	0.68	0.67	0.23	0.20	0.26
Paternal relationships ^a	0.52	0.50	0.50	0.22	0.22	0.25
Balanced cohesion ^b	–	0.56	0.59	–	0.43	0.43
Flexibility ^b	–	0.57	0.56	–	0.39	0.41

^a Study 1 sample, child report

^b Study 2 sample, parent report; all correlations (*r*), *p* < 0.01

metric. As a result, we did not eliminate any child-report items because of DIF. For parent-report, we observed no statistically significant DIF due to gender or ethnicity. And, like child-report, though we observed statistically significant DIF due to race and due to age, this DIF did not appear to have a substantive impact on scores. As with child report, scores ignoring and adjusting for DIF were nearly identical. As a result, we did not eliminate any proxy-report items because of DIF.

Scoring procedures for the SF-8 and SF-4 are presented in Online Resource 8. Child-report family relationships scores were strongly positively correlated with Emotional Support, and the correlation with Maternal Relationships was larger than that with Paternal Relationships. As hypothesized, parent-reported family relationships were more highly correlated with Balanced Cohesion than Flexibility scores (Table 3). Known-group comparisons revealed that family relationships (by both child- and parent-report) were weaker among adolescents than younger children, as expected (Table 4). African-American and Asian parents reported that their children had better family relationships, compared with White children, although this difference was not reported by children. Children classified as “other” race had lower self-reported Family Relationships scores. Hispanic children reported poorer family relationships than their non-Hispanic counterparts, a

difference not reported by parents. Finally, children with chronic conditions had poorer parent-reported family relationships.

Discussion

Research to explain the complex role of families in child and adolescent health development as well as their role in the management of chronic conditions in youth is hampered by a lack of well-validated measures of children’s family relationship experiences. The PROMIS Pediatric Family Relationships instruments were developed to fill this critical gap. The Family Relationships item banks and short forms were developed using a rigorous mixed-method approach consistent with PROMIS standards [8, 9]. The resulting instruments are conceptually grounded in theories of attachment, the proximal processes posited by bio-ecological theory, and research on the impact of families on psychobiological development [15–17]. This conceptual foundation was refined by expert researchers and child and parent perspectives and provided a foundation for the instruments’ content validity. Items were informed by a systematic review of existing measures, written according to PROMIS standards, and iteratively revised through translatability reviews and cognitive interviews that

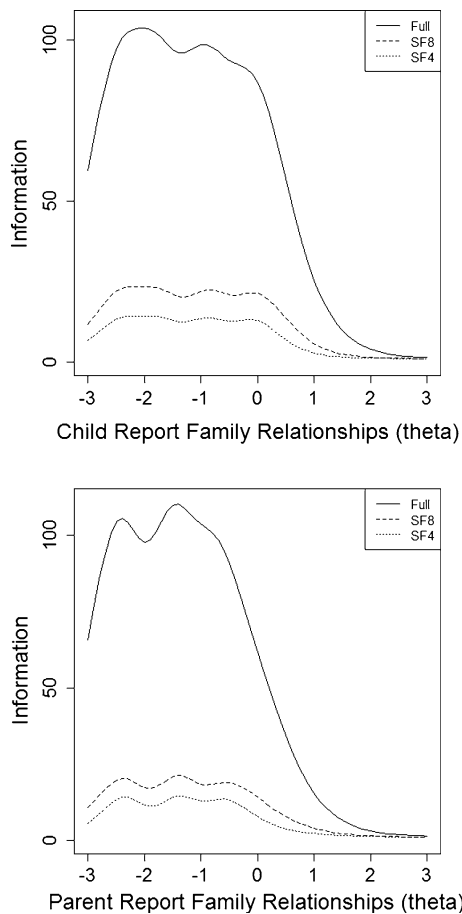


Fig. 2 Test information functions (TIFs): PROMIS pediatric family relationships instruments. These figures show the amount of information, which is similar to precision and reliability, across levels of the latent variable (theta)

Table 4 Known-group comparisons (Cohen's *d* effect size estimates)

	Child-report			Parent-report		
	Full bank	SF-8	SF-4	Full bank	SF-8	SF-4
Age (ref = 8–12 years)						
13–17 years	-0.32	-0.28	-0.21	-0.28	-0.21	-0.26
Gender (ref = girls)						
Boys	0.02	0.05	0.05	0.08	0.10	0.07
Race (ref = White)						
African American/Black	-0.11	-0.06	-0.07	0.25	0.35	0.31
Asian	-0.08	-0.02	-0.04	0.24	0.39	0.37
Other	-0.29	-0.29	-0.27	-0.18	-0.15	-0.21
Ethnicity (ref = Non-Hispanic)						
Hispanic	-0.28	-0.22	-0.18	-0.09	-0.05	-0.07
Child chronic condition ^a (ref = no condition)						
Has a condition	-0.09	-0.09	-0.12	-0.23	-0.20	-0.22

Analyses conducted with child- ($N = 1845$) and parent-report ($N = 960$) data collected in study 1; d = Cohen's d effect size statistic; Bold indicates significant effect sizes ($p < 0.001$); *ref* = referent group

^a Parent-report variable

maximized the tools' developmental, linguistic, and cultural appropriateness.

Psychometric analyses of data collected from over 5000 children and parents were performed to generate item banks and short forms. Two aspects of family function, organization and predictability were not retained because they were sufficiently distinct from the concept of relationship quality. The resulting instruments assess a single Family Relationships domain, children's subjective experiences of being supported by one's family, feeling accepted, valued, loved and cared for, feeling that they can communicate with family and that family members, especially parents, can be trusted and depended on for help and understanding. They measure the degree to which the core human needs for security and value are met, which are essential to children's ability to explore opportunities and take developmentally appropriate risks [1, 3, 21].

The item banks and short forms provide precise and reliable measurement across a wide range of quality in family relationships irrespective of child age, gender, race/ethnicity, and chronic disease status. Known-group comparisons and convergence with legacy measures provided preliminary evidence of construct validity. Multi-group IRT was used to calibrate item banks such that scores estimated using the IRT parameters can be compared to a US population mean of 50 and standard deviation of 10. IRT-based calibration enables CAT administration of the item banks, which maintains measurement precision based on a smaller number of completed items. CAT versions of the instruments can be administered using the PROMIS Assessment Center (www.healthmeasures.net).

All final Family Relationship items are positively oriented, yet as shown in the test information functions, they provide relatively more information about children with average to below-average family relationship quality. Although this is also true for most existing family assessment tools, the unique positive orientation of the PROMIS Family Relationships measures may have greater potential to detect improvements in family relationships over time.

Limitations

Despite the intensive application of state of the art instrument development methods, several limitations are worth noting. First, the child and parent concept elicitation interviews were conducted in one setting in which most participants were non-Hispanic Black or White. Children of Hispanic descent were not involved until the calibration survey phase, and other culturally distinct subgroups are not well represented. However, the items were carefully reviewed for culturally meaningful translation. Second, although children with chronic medical conditions were purposively recruited into the qualitative phase and the calibration samples, all were sufficiently healthy to attend outpatient clinics. Thus, the samples may underrepresent children with more severe conditions and disabilities. Also, the single parent-report item used to identify children with chronic conditions in Study 1 and 2 did not identify the child's condition, preventing more in-depth analysis of disorders and family relationships. Third, data used for instrument calibration were collected using a variety of modalities: in person, at school, and via Internet. Research conducted with adults suggests that questionnaires self-administered online produce scores that are generally equivalent to those derived from self- or proctor-administered paper questionnaires [49]. However, research on mode effects for children is scarce. The degree to which assessment modality impacted the instruments' measurement properties is unknown. Fourth, although we were able to test and confirm the instrument dimensionality across two different samples, future research should replicate our dimensionality analyses. Single-factor CFA models suggest that family organization and predictability can be differentiated from children's perceptions of involvement, care, support, and understanding within their families. The Family Relationships measures were developed in accordance with the PROMIS analytic strategy of generating a large item bank that measures a single common dimension [38]. However, the excluded items provide valuable information about unique and important aspects of family functioning. As item banking initiatives advance toward multidimensionality and the application of bi-factor models [50–52], the utility of forming Family Relationships

subscales should be reconsidered. Lastly, additional validation of the measures in socio-demographic and clinical subgroups is warranted, particularly evaluations of the instruments' sensitivity to change.

Conclusion

The newly developed PROMIS Pediatric Family Relationships item banks and short forms provide meaningful and psychometrically robust measures of children's subjective family relationship experiences. The tools are appropriate for use with children aged 8–17 and parents of children aged 5–17 years. They can support research that evaluates how children's subjective family relationship experiences are established, and impact their health, development, and quality of life. Findings from such studies, especially in longitudinal research, can guide program and policy interventions to enhance family relationships, and thereby improve health across the life course.

Funding PROMIS[®] was funded with cooperative agreements from the National Institutes of Health (NIH) Common Fund Initiative (Northwestern University, PI: David Cella, PhD, U54AR057951, U01AR052177; Northwestern University, PI: Richard C. Gershon, PhD, U54AR057943; American Institutes for Research, PI: Susan (San) D. Keller, PhD, U54AR057926; State University of New York, Stony Brook, PIs: Joan E. Broderick, PhD and Arthur A. Stone, PhD, U01AR057948, U01AR052170; University of Washington, Seattle, PIs: Heidi M. Crane, MD, MPH, Paul K. Crane, MD, MPH, and Donald L. Patrick, PhD, U01AR057954; University of Washington, Seattle, PI: Dagmar Amtmann, PhD, U01AR052171; University of North Carolina, Chapel Hill, PI: Harry A. Guess, MD, PhD (deceased), Darren A. DeWalt, MD, MPH, U01AR052181; Children's Hospital of Philadelphia, PI: Christopher B. Forrest, MD, PhD, U01AR057956; Stanford University, PI: James F. Fries, MD, U01AR052158; Boston University, PIs: Alan Jette, PT, PhD, Stephen M. Haley, PhD (deceased), and David Scott Tulsky, PhD (University of Michigan, Ann Arbor), U01AR057929; University of California, Los Angeles, PIs: Dinesh Khanna, MD (University of Michigan, Ann Arbor) and Brennan Spiegel, MD, MSHS, U01AR057936; University of Pittsburgh, PI: Paul A. Pilkonis, PhD, U01AR052155; Georgetown University, PIs: Carol M. Moinpour, PhD (Fred Hutchinson Cancer Research Center, Seattle) and Arnold L. Potosky, PhD, U01AR057971; Children's Hospital Medical Center, Cincinnati, PI: Esi M. Morgan DeWitt, MD, MSCE, U01AR057940; University of Maryland, Baltimore, PI: Lisa M. Shulman, MD, U01AR057967; and Duke University, PI: Kevin P. Weinfurt, PhD, U01AR052186). NIH Science Officers on this project have included Deborah Ader, PhD, Vanessa Ameen, MD (deceased), Susan Czajkowski, PhD, Basil Eldadah, MD, PhD, Lawrence Fine, MD, DrPH, Lawrence Fox, MD, PhD, Lynne Haverkos, MD, MPH, Thomas Hilton, PhD, Laura Lee Johnson, PhD, Michael Kozak, PhD, Peter Lyster, PhD, Donald Mattison, MD, Claudia Moy, PhD, Louis Quatrano, PhD, Bryce Reeve, PhD, William Riley, PhD, Peter Scheidt, MD, Ashley Wilder Smith, PhD, MPH, Susana Serrate-Sztejn, MD, William Phillip Tonkins, DrPH, Ellen Werner, PhD, Tisha Wiley, PhD, and James

Witter, MD, PhD. The contents of this article use data developed under PROMIS. These contents do not necessarily represent an endorsement by the US Federal Government or PROMIS. See www.nihpromis.org for additional information on the PROMIS® initiative.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497–529. doi:10.1037/0033-2909.117.3.497.
- Fiese, B. H., Winter, M., Anbar, R., Howell, K., & Poltrock, S. (2008). Family climate of routine asthma care: Associating perceived burden and mother-child interaction patterns to child well-being. *Family Process*, *47*, 63–79. doi:10.1111/j.1545-5300.2008.00239.x.
- Maslow, A. H. (1968). *Toward a psychology of being*. New York: Van Nostrand.
- Olson, D. (2011). FACES-IV and the circumplex model: Validation study. *Journal of Marital and Family Therapy*, *37*, 64–80. doi:10.1111/j.1752-0606.2009.00175.x.
- Moos, R. H., & Moos, B. S. (1994). *Family environment scale manual*. Palo Alto, CA: Consulting Psychologists Press.
- Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., et al. (2007). The patient-reported outcomes measurement information system (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Medical Care*, *45*(5 Suppl 1), S3–S11. doi:10.1097/01.mlr.0000258615.42478.55.
- Forrest, C. B., Bevans, K. B., Tucker, C., Riley, A. W., Ravens-Sieberer, U., Gardner, W., et al. (2012). The patient reported outcome measurement information system (PROMIS) for children and youth: Application to pediatric psychology. *Journal of Pediatric Psychology*, *37*(6), 614–621.
- DeWalt, D. A., Rothrock, N., Yount, S., & Stone, A. A. (2007). Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care*, *5*(1), S12–S21.
- Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement information system (PROMIS). *Medical Care*, *45*(5 Suppl 1), S22–S31.
- PROMIS. PROMIS instrument development and psychometric evaluation scientific standards. Retrieved May 21, 2016, from http://www.nihpromis.org/Documents/PROMIS_Standards_050212.pdf.
- Bjorner, J. B., Chang, C. H., Thissen, D., & Reeve, B. B. (2007). Developing tailored instruments: item banking and computerized adaptive assessment. *Quality of Life Research*, *16*(Suppl 1), 95–108.
- Teresi, J. A., & Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research*, *16*(Suppl 1), 33–42.
- van der Linden, W. J., & Hambleton, R. K. (1997). *Item response theory: Brief history, common models, and extensions* (pp. 1–28). New York: Springer.
- Hahn, E. A., DeVellis, R. F., Bode, R. K., Garcia, S. F., Castel, L. D., Eisen, S. V., et al. (2010). Measuring social health in the patient-reported outcomes measurement information system (PROMIS): Item bank development and testing. *Quality of Life Research*, *19*, 1035–1044. doi:10.1007/s11136-010-9654-0.
- Bowlby, J. (1973). *Attachment and Loss: Vol. 2. separation*. New York: Basic Books.
- Gunnar, M. R. (1998). Quality of early care and buffering of neuroendocrine stress reactions: Potential effects on the developing human brain. *Preventive Medicine*, *27*, 208–211. doi:10.1006/pmed.1998.0276.
- Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of social development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Vol 1, theoretical models of human development* (6th ed., pp. 993–1023). New York: Wiley.
- Forrest, C. B. (2014). A living systems perspective on health. *Medical Hypotheses*, *82*, 209–214. doi:10.1016/j.mehy.2013.11.040.
- Diamond, L. M., & Fagundes, C. P. (2008). Developmental perspectives on links between attachment and affect regulation over the lifespan. *Advances in Child Behavior and Development*, *36*, 83–134.
- Sroufe, L. A. (2005). Attachment and development: A prospective, longitudinal study from birth to adulthood. *Attachment & Development*, *7*, 349–367. doi:10.1080/14616730500365928.
- Sroufe, L. A., Coffino, B., & Carlson, E. A. (2010). Conceptualizing the role of early experience: Lessons from the Minnesota Longitudinal Study. *Developmental Review*, *30*, 36–51. doi:10.1016/j.dr.2009.12.002.
- Chen, E., & Schreier, H. M. (2008). Does the social environment contribute to asthma? *Immunology and Allergy Clinics of North America*, *28*(3), 649–664. doi:10.1016/j.iac.2008.03.007.
- Davies, P., & Woitach, M. (2008). Children's emotional security in the interparental relationship. *Current Directions in Psychological Science*, *17*, 269–274. doi:10.1111/j.1467-8721.2008.00588.x.
- Wood, B. L., Klebba, K. B., & Miller, B. D. (2000). Evolving the biobehavioral family model: The fit of attachment. *Family Process*, *39*, 319–344. doi:10.1111/j.1545-5300.2000.39305.x.
- Cox, M. J., & Paley, B. (1997). Families as systems. *Annual Review of Psychology*, *48*, 243–267. doi:10.1146/annurev.psych.48.1.243.
- Umberson, D., & Montez, J. K. (2010). Social relationships and health: a flashpoint for health policy. *Journal of Health and Social Behavior*, *51*, S54–S66. doi:10.1177/0022146510383501.
- Fiese, B. H., & Wamboldt, F. (2000). Family routines, rituals, and asthma management: A proposal for family-based strategies to increase treatment adherence. *Family Systems and Health*, *18*, 405–418. doi:10.1037/h0091864.
- Lasch, K. E., Marquis, P., Vigneux, M., Abetz, L., Arnould, B., Bayliss, M., et al. (2010). PRO development: rigorous qualitative research as the crucial foundation. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *19*(8), 1087–1096. doi:10.1007/s11136-010-9677-6.
- Klem, M., Saghabi, E., Abromitis, R., Stover, A., Dew, M. A., & Pilkonis, P. (2009). Building PROMIS item banks: Librarians as co-investigators. *Quality of Life Research*, *18*(7), 881–888. doi:10.1007/s11136-009-9498-7.
- Irwin, D. E., Varni, J. W., Yeatts, K., & Dewalt, D. A. (2009). Cognitive interviewing methodology in the development of a pediatric item bank: A patient reported outcomes measurement information system (PROMIS) study. *Health and Quality of Life Outcomes*, *7*(1), 3.
- Wild, D., Eremenco, S., Mear, I., Martin, M., Houchin, C., & Gawlicki, M. (2009). Multinational trials—recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: The ISPOR patient-reported outcomes translation and linguistic validation good. *Value in Health*, *12*(4), 430–440.

32. Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., & Verjee-Lorenz, A. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health*, 8(2), 94–104.
33. IQOLA. (2011). The international quality of life assessment (IQOL) project.
34. Salsman, J. M., Butt, Z., Pilkonis, P., Cyranowski, J. M., Zill, N., Hendrie, H. C., et al. (2013). Emotion assessment using the NIH Toolbox. *Neurology*, 80(11 Supplement 3), S45–S48. doi:10.1212/WNL.0b013e3182872dd2.
35. DiSogra, C., Dennis, J., & Fahimi, M. (2010). On the quality of ancillary data available for address-based sampling. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 4174–4183.
36. Dennis, J. (2010). Knowledgepanel®: Processes & procedures contributing to sample representativeness & tests for self-selection Bias. Retrieved June 6, 2016, from <http://www.knowledge-networks.com/ganp/docs/KnowledgePanelR-Statistical-Methods-Note.pdf>.
37. Lohr, S. (2009). *Sampling: design and analysis*. Toronto, ON: Nelson Education.
38. Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement information system (PROMIS). *Medical Care*, 45(5 Suppl 1), S22–S31. doi:10.1097/01.mlr.0000250483.85507.04.
39. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
40. Bollen, K. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
41. Carle, A. (2010). Mitigating systematic measurement error in comparative effectiveness research in heterogeneous populations. *Medical Care*, 48(6), S68.
42. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. doi:10.2307/2346101.
43. Choi, S., Gibbons, L., & Crane, P. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39(8), 1–30.
44. Millsap, R., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, 39(3), 479–513.
45. Green, B. F., Bock, D., Humphres, R. L., & Linn, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21(4), 347–360.
46. Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431–444.
47. Choi, S. (2009). Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. *Applied Psychological Measurement*, 33(8), 644–645.
48. Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
49. Gundy, C. M., & Aaronson, N. K. (2010). Effects of mode of administration (MOA) on the measurement properties of the EORTC QLQ-C30: A randomized study. *Health and Quality of Life Outcomes*, 8, 35–42.
50. Croudace, T. J., & Jan, R. B. (2014). Item bank measurement of depression: will one dimension work? *Journal of Clinical Epidemiology*, 67, 4–6. doi:10.1016/j.jclinepi.2013.08.002.
51. Reise, S., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16, 19–31.
52. Stochl, J., Bohnke, J. R., Pickett, K. E., & Croudace, T. J. (2016). Computerized adaptive testing of population psychological distress: simulation-based evaluation of GHQ-30. *Social Psychiatry and Psychiatric Epidemiology*, 51, 895–906. doi:10.1007/s00127-015-1157-4.