

Psychometric Evaluation of the PROMIS[®] Pediatric Psychological and Physical Stress Experiences Measures

Katherine B. Bevans,¹ PhD, William Gardner,^{2,3} PhD, Kathleen A. Pajer,² MD, Brandon Becker,⁴ PhD, Adam Carle,⁵ PhD, Carole A. Tucker,¹ PhD, and Christopher B. Forrest,^{4,6} MD, PhD

¹Temple University, ²University of Ottawa, ³Children's Hospital of Eastern Ontario, ⁴The Children's Hospital of Philadelphia, ⁵Cincinnati Children's Hospital, and ⁶University of Pennsylvania Perelman School of Medicine

All correspondence concerning this article should be addressed to Katherine Bevans, PhD, Department of Rehabilitation Sciences, Temple University, Philadelphia, PA 19121, USA. E-mail: katherine.bevans@temple.edu

Received July 6, 2017; revisions received February 3, 2018; accepted January 17, 2018

Abstract

Objective To provide psychometric evaluation of the PROMIS[®] Pediatric Psychological and Physical Stress Experiences measures. **Methods** Across two studies, Psychological and Physical Stress Experiences items were administered to 2,875 children aged 8–17 years and 2,212 parents of children aged 5–17 years. Analyses included descriptive statistics, reliability, factor analysis, differential item functioning (DIF), and assessment of construct validity. Items were calibrated using item response theory to estimate item parameters representative of the United States. Recommended eight- and four-item short forms were constructed for child- and parent-report versions of the Psychological and Physical Stress Experiences item banks. **Results** Final item banks were unidimensional and items were locally independent and free from impactful DIF. Psychological Stress banks include 19 child-report and 12 parent-proxy items. Physical Stress banks include 26 child-report and 26 parent-proxy items. All instruments have strong internal consistency and retest-reliability, and provide precise estimates of varying stress levels. The instruments' construct validity was evidenced by known-group comparisons and convergence with legacy measures. **Conclusions** The Patient Reported Outcome Measurement Information System (PROMIS) Pediatric Psychological and Physical Stress item banks and short forms provide efficient, precise, and valid assessments of children's stress experiences.

Key words: assessment; children; stress.

Children who endure frequent or severe stress exposure are at increased risk for poorer physical, cognitive, social, and emotional functioning (Kugler, Bloom, Kaercher, Truax, & Storch, 2012; Shonkoff et al., 2012; Watt, Weber, Davies, & Forster, 2017). Stress is a dynamic transactional process involving stimuli exposure (e.g., the stressor), detection and appraisal of the stimuli, and responses to those stimuli that are perceived as threatening (Gould, Hussong, & Keeley, 2008). Whereas mild or moderate stress exposure tends to enhance motivation and productivity,

persistent and high-intensity stress can lead to physical changes, negative self-evaluation, pessimism, hopelessness, anxiety, depression, and acting out (Kugler et al., 2012). Severe and frequent stress may compromise attention, memory, lead to confusion, and reduce one's capacity to effectively resolve or adapt to stressors, thereby further magnifying the negative effects of stress (Garmezy, 1986). Over time, repetitive exposure to stress and its negative correlates may activate physiological responses or prompt behaviors that increase risk of infectious illness (Graham, Douglas, & Ryan,

1986), injury (Boyce et al., 1995), and psychiatric disorders (Rutter, 1989).

Some physiological responses to stressors, such as temperature regulation, are biological adaptations that are not recognized by the individual. Others are knowingly experienced. Stress is *experienced* when a person perceives challenges that are threatening and exceed his or her capacity to adapt to them (Cohen, Kessler, & Gordon, 1997; Kupst et al., 2015). Physical stress experiences are the *consciously perceived manifestations of the body's fight-or-flight response to threat, mediated by the sympathetic nervous system, and its impact on the cardiovascular, respiratory, musculoskeletal, and gastrointestinal systems*. These experiences include physical arousal (e.g., sensory alertness, muscle potential), agitation (e.g., restlessness, fidgetiness), pain, sleep disturbance, and gastrointestinal distress (Laurent, Catanzaro, & Joiner, 2004). Psychological stress experiences are *thoughts and feelings that occur in response to threatening events*. Commonly referred to as “feeling stressed out,” these experiences include believing that events are uncontrollable, unpredictable, or overloading; disruptions to perception and cognition; and feeling nervous, worried, and irritable (Cohen et al., 1997).

If heightened stress experiences are detected early, interventions that lessen stress exposure and enhance coping may reduce symptoms and prevent the development of stress-related conditions (Franke, 2014; Kugler et al., 2012). For example, cognitive behavioral interventions reduce anxiety and depression by teaching children and their parents to identify bodily and emotional stress responses and challenge the beliefs that elicit those experiences (Storch et al., 2008). Mindfulness-based interventions, particularly those focused on awareness of psychological and physical manifestations of stress, reduce some children's anxiety and emotional and behavioral reactivity to stress (Burke, 2010; Harnett & Dawe, 2012). Recognizing and quantifying children's stress experiences requires the use of sensitive, developmentally appropriate, reliable, and accurate measurement tools. Because most stress experiences are internally perceived, they are best assessed by self-report.

The Perceived Stress Scale (PSS) is the most frequently used self-report measure of psychological stress experiences (Cohen & Williamson, 1988). Although the PSS was originally developed for adults, the National Institutes of Health Toolbox for the Assessment of Neurological and Behavioral Function initiative recommends using a modified (10-item) version of the measure with children aged 12–17 years (self-report) and parents of children aged 8–12 years (proxy-report) (Kupst et al., 2015). The Perceived Stress Scale-10 (PSS10) is not recommended for use with children aged <12 years. Items contain terms

that are difficult for younger children (e.g., unexpectedly, effectively, irritations, overcome) and items have a Flesch–Kincaid grade-level equivalent of 6.5 (range: 2.8–10.4; $SD = 2.7$) (Klare, 1976). Item understandability is especially important for child-report measures because the competencies needed to reliably and accurately self-report are acquired during a child's development (Bevans, Moon, Riley, & Forrest, 2010).

Several commonly used child-report anxiety measures including the Revised Children's Manifest Anxiety Scale, and the Multidimensional Anxiety Scale for Children have physical stress subscales (March, Parker, Sullivan, Stallings, & Conners, 1997; Reynolds & Paget, 1981). These subscales do not exhaustively cover children's physical stress experiences and many conflate physically experienced stress with cognitive and affective stress responses (e.g., rumination, fear) (Laurent et al., 2004). A few child-report instruments were developed specifically to measure child's physical symptoms, especially bodily manifestations of autonomic arousal. These include the Children's Somatization Inventory (CSI) (Walker, Beck, Garber, & Lambert, 2009; Walker & Garber, 1992) and the Physiological Hyperarousal Scale for Children (Laurent et al., 2004). Some instruments were developed for specific clinical subgroups, and their generalizability to other populations has not been evaluated. For example, the Children's Psychosomatic Symptom Checklist was developed and validated for children with chronic abdominal pain (Walker et al., 2009). Other measures lack evidence of content validity for pediatric populations in general. The Children's Psychosomatic Symptom Checklist was developed by refining the adult Psychosomatic Symptom Checklist to enhance its suitability for children, but the instrument's relevance, meaningfulness, and understandability to children has not been evaluated using best-practice qualitative methods (Lasch et al., 2010).

We developed the Patient Reported Outcome Measurement Information System (PROMIS®) Pediatric Psychological and Physical Stress Experiences child- and parent-proxy measures to address the limitations of existing child-report measures of stress experiences. The psychological stress item pool was developed to assess child and parent perspectives on children's cognitive-perceptual disruption, perceptions of controllability and manageability, anger, and fear. The physical stress item pool was designed to measure children's arousal, agitation, pain, sleep disturbance, and gastrointestinal distress by child- and parent-report (Bevans et al., 2013).

Consistent with the PROMIS approach, the Psychological and Physical Stress measures use a domain-specific rather than disease-specific measurement approach (Magasi, Ryan, & Revicki, 2012). Domains are clinically coherent and empirically unidimensional health attributes. The domain-specific

approach assumes that health attributes are not unique to a specific health condition. However, disorders may have characteristic profiles within domains. Domain-specific measures enable comparisons across conditions and support pragmatic research involving people with multiple conditions (Broderick, Morgan-DeWitt, Rothrock, Crane, & Forrest, 2013).

As with all PROMIS measures, initial Psychological and Physical Stress item pools were developed in accordance with rigorous content validation standards (Matza et al., 2013). Previously, we described the qualitative procedures used to ensure that the tools are conceptually grounded in life course models of health, represent child and parent perspectives, and contain items that are well understood by the intended respondents, culturally harmonized, and linguistically translatable (Bevans, Gardner, Pajer, Riley, & Forrest, 2013). On completion of the qualitative instrument development phase, the Psychological and Physical Stress Experiences item pools included 63 and 40 items, respectively. Here, we present results of two studies that were done to produce final versions of the instruments and evaluate their psychometric properties. In Study 1, child and parent responses to 63 psychological and 40 physical stress items were subjected to psychometric analyses and item response theory (IRT)-based item calibration. The purpose of Study 1 was to finalize the item banks and short-form versions of the measures. In Study 2, the IRT-based item parameters were adjusted to better represent the U.S. general population. Both studies evaluated the instruments' construct validity.

Method

All study procedures were approved by the institutional review board at the Children's Hospital of Philadelphia (CHOP). Parent's informed consent and child's assent were obtained for all youth participants. Parents consented for their own participation.

Participants and Procedures

Study 1

Study 1 was conducted to identify items for inclusion in the final child- and parent-report Psychological and Physical Stress item banks. Participants were 1,874 children aged 8–17 years and 924 parents of children aged 5–17 years. Thirty Study 1 parent participants had 5–7-year-old children. Parent-report data were collected for 894 of the 1,874 children who participated in Study 1 (47.7%). To ensure the measures' applicability to heterogeneous populations, Study 1 participants were purposively sampled to maximize representation of sociodemographically diverse families and children with a range of health-related concerns. The number of potential participants

approached and those who refused to participate were not tracked in Study 1.

About half of Study 1 participants (55% of children; 53% of parents) were recruited from an opt-in panel (Op4G), a private research community with approximately 250,000 members who participate in Internet-based research activities from their home computers. Op4G maintains a national sample of participants whose demographic information is updated on a regular basis (see Op4G.com). Adult members of the Op4G panel known to have children aged 5–17 years were notified by e-mail of their eligibility to participate in the Study. Parents who consented were e-mailed a link to the online parent questionnaire. After completing their measures, parents of children aged 8–17 years asked their child to complete the child-report instruments. A randomly selected group of participants from the opt-in panel were recontacted and reassessed 2 weeks later ($M = 22.8$ days, $SD = 6.5$ days) to evaluate test–retest reliability of the measures.

As with many opt-in panels, Hispanic, African-American, and lower-income families are underrepresented in the Op4G sample. Therefore, we recruited additional participants (children: 37%, parents: 28%) from school districts in New Hampshire, Vermont, and Texas that were diverse with respect to race, ethnicity, socioeconomic status, and geographic location. All English-speaking students in Grades 3–12, except those in self-contained special education classrooms, were invited to participate in the study. Children with parental consent were administered questionnaires at school in groups of 25–150 students monitored by school staff and research assistants. Parent questionnaires were sent home with students with postage-paid envelopes that parents used to return them directly to the investigators.

Finally, to ensure inclusion of children with medical concerns or chronic health conditions, we recruited participants from CHOP clinics (children: 9%, parents: 19%). Child sample sizes from the clinics were primary care ($n = 34$), emergency department ($n = 90$), gastroenterology ($n = 24$), rheumatology ($n = 10$), and healthy weight ($n = 3$). Parents and children who met study inclusion criteria (child age and capacity to self-report) were contacted by study staff, consented and assented if interested, and given the option to complete the study materials at home or during their medical visit using the Internet-based NIH Assessment Center data collection platform (www.healthmeasures.net).

Study 1 participants completed the Psychological and Physical Stress item banks. They were also administered measures of related constructs to enable the assessment of convergent validity. Given the large number of items, it was not feasible to administer all measures to the same participants. To reduce participant burden, a randomly selected subset of

child-parent dyads completed the PSS10. Two additional unique subgroups of children completed PROMIS measures. One group completed the PROMIS Pediatric Depression, Anxiety, and Anger short forms. The other group completed the PROMIS Fatigue and Pain Interference short forms. Study 1 sample sizes are shown in [Tables I](#).

Study 2

In Study 2, population-representative data were used to adjust item parameters, so that Psychological and Physical Stress scores can be interpreted relative to the noninstitutionalized U.S. population of children aged 5–17 years. Study 2 participants were recruited from GfK Knowledge panel, a dual frame (random-digit dial and address-based) online panel of a representative random sample of the U.S. population ([Dennis, 2010](#); [DiSogra, Dennis, & Fahimi, 2010](#)). A total of 53,340 GfK panelists known to have children aged 5–17 years were notified by e-mail of their eligibility to participate. Of those, 15,114 parents accepted the invitation to learn more about the study and 14,974 consented to participate. A total of 10,708 parents were deemed eligible for the study. The majority of ineligible parents were excluded because quotas for their child's age and gender had already been met or because the child had a cognitive limitation that prevented him/her from participating. A subset of eligible parents ($n = 1,606$) was randomly selected to complete the stress measures; remaining parents provided data that were used to evaluate other PROMIS measures not described here. Of these, 1,314 parents of children aged 5–17 years and 1,014 children aged 8–17 years completed the stress measures. All children in Study 2 had parent-report data and 300 parent participants had children aged 5–7 years (who did not self-report). In both studies, parents were restricted to completing measures for a single child.

Parents who consented were e-mailed a link to the online parent questionnaire. After completing their measures, parents of children aged 8–17 years asked their child to complete the child questionnaires. Data collection continued until age and gender quotas were met for each form. The panel data were weighted, so that 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) ([Lohr, 2009](#)). Two unique randomly selected subgroups of children completed validation measures. One group completed the PSS10 and the other group completed the Children's Somatization Inventory-24 (CSI-24). Study 2 sample sizes are shown in [Table I](#).

Measures

Demographics

Child- and parent-reports of their own and their child's gender, age, race, and ethnicity were used in the analysis of child-report and parent-proxy measures, respectively. Parents reported their educational attainment and annual household income.

Special Health-Care Needs

In Study 2, parents completed the Children with Special Health Care Needs Screener (CSHCN), a measure of chronic, behavioral, and neurodevelopmental health problems that require health services or cause functional limitations ([Bethell et al., 2002](#)). The CSHCN screen is positive if the child has a condition lasting at least 12 months, and the parent-reports that the child has any of the following qualifying indicators: (1) needs or uses medicine prescribed by a doctor, other than vitamins; (2) needs or uses more medical care, mental health, or educational services than is usual for most children of the same age; (3) is limited in his or her ability to do the things most children of the same age can do; (4) needs or gets special therapy such as physical, occupational, or speech therapy; or, (5) has an emotional, developmental, or behavioral problem for which he or she needs treatment or counseling.

PROMIS Pediatric Psychological and Physical Stress Measures

Study 1 participants answered 63 psychological and 40 physical stress items using five-point Likert response categories that ranged from 1 (*Never*) to 5 (*Always*) ([Bevans et al., 2013](#)). Positively worded items were reverse-scored, so that for all items, higher scores indicated greater stress. Study 2 participants completed eight psychological and eight physical stress items that were selected for eight-item short-form versions of the measures.

Perceived Stress Scale-10

Children aged 8–17 years completed the PSS10, a self-report measure that assesses the degree to which life situations are appraised as stressful over the past month ([Cohen & Williamson, 1988](#)). Parents of children aged 5–17 years gave their impressions of their children's stress levels by completing the PSS10 by proxy report. An abbreviated (10-item) version of the measure (PSS10) has adequate reliability for 13–17 year-olds ($\alpha = 0.89$) and for parents of 8–12 year-olds ($\alpha = 0.87$) ([Kupst et al., 2015](#)). In this study, the PSS10 had adequate internal consistency reliability by child-report ($\alpha = 0.86$) and parent-report ($\alpha = 0.87$).

Table 1. Sample Sociodemographic Characteristics, N (%)

Participant characteristic	Study 1						Study 2						
	Psychological stress			Physical stress			Psychological stress			Physical stress			
	Child N = 1,874	Parent N = 913	Child N = 1,843	Parent N = 924	Child N = 1,001	Parent N = 1,299	Child N = 1,013	Parent N = 1,313	Child N = 1,013	Parent N = 1,313			
Subsamples													
Internet	1,024 (54.6)	482 (52.8)	1015 (55.1)	480 (52.0)	1,001 (100.0)	1,299 (100.0)	1,013 (100.0)	1,313 (100.0)	1,013 (100.0)	1,313 (100.0)			
School	689 (36.8)	255 (27.9)	677 (36.7)	271 (29.3)	-	-	-	-	-	-			
Clinic	161 (8.6)	176 (19.3)	151 (8.2)	173 (18.7)	-	-	-	-	-	-			
Child age (years)													
5-7	-	30 (3.3)	-	30 (3.3)	-	298 (22.9)	-	-	-	300 (22.9)			
8-12	896 (47.8)	424 (46.4)	887 (48.1)	428 (46.3)	502 (50.1)	502 (38.7)	501 (49.5)	501 (38.2)	501 (49.5)	501 (38.2)			
13-17	978 (52.2)	459 (50.3)	956 (51.9)	466 (50.4)	499 (49.9)	499 (38.4)	512 (50.5)	512 (39.0)	512 (50.5)	512 (39.0)			
Child gender, male	972 (51.9)	459 (50.3)	955 (51.9)	463 (50.2)	494 (51.0)	642 (51.0)	516 (51.0)	666 (51.0)	516 (51.0)	666 (51.0)			
Child race													
White	1,312 (71.3)	646 (72.9)	1,297 (71.6)	656 (73.1)	734 (70.5)	955 (69.7)	797 (69.5)	1,015 (68.9)	797 (69.5)	1,015 (68.9)			
African-American/Black	247 (13.4)	134 (15.1)	237 (13.1)	135 (15.1)	96 (14.3)	128 (14.4)	82 (14.9)	114 (15.1)	82 (14.9)	114 (15.1)			
Asian	82 (4.5)	42 (4.7)	79 (4.4)	40 (4.5)	46 (2.6)	56 (2.9)	34 (3.8)	42 (3.5)	34 (3.8)	42 (3.5)			
Other	199 (10.8)	64 (7.2)	199 (11.0)	66 (7.4)	125 (12.6)	160 (13.0)	98 (11.9)	140 (12.5)	98 (11.9)	140 (12.5)			
Child ethnicity, Hispanic	283 (15.2)	89 (10.0)	279 (15.2)	89 (9.9)	156 (22.8)	216 (23.3)	150 (22.7)	213 (23.2)	150 (22.7)	213 (23.2)			
Annual household income ^a													
<\$40,000	-	-	-	-	253 (32.2)	351 (32.7)	242 (31.5)	344 (32.6)	242 (31.5)	344 (32.6)			
≥\$40,000e	-	-	-	-	748 (67.9)	948 (67.3)	771 (68.5)	969 (67.4)	771 (68.5)	969 (67.4)			
Parent educational attainment ^a													
High school diploma, equivalence, or less	-	-	-	-	174 (33.7)	230 (33.4)	167 (31.2)	210 (29.6)	167 (31.2)	210 (29.6)			
Some college/associate's degree	-	-	-	-	364 (33.0)	460 (31.8)	355 (34.3)	476 (35.4)	355 (34.3)	476 (35.4)			
College degree or higher	-	-	-	-	463 (43.3)	609 (34.8)	491 (34.6)	627 (35.0)	491 (34.6)	627 (35.0)			
Child special health-care need	-	-	-	-	193 (19.4)	232 (18.0)	195 (19.5)	235 (18.1)	195 (19.5)	235 (18.1)			
Needs/uses medication (prescribed)	-	-	-	-	110 (11.1)	132 (10.2)	96 (9.6)	117 (9.0)	96 (9.6)	117 (9.0)			
Needs/uses medical, mental health, or psychological services	-	-	-	-	64 (6.4)	81 (6.3)	62 (6.2)	73 (5.6)	62 (6.2)	73 (5.6)			
Has functional limitation	-	-	-	-	43 (4.3)	57 (4.4)	45 (4.4)	59 (4.5)	45 (4.4)	59 (4.5)			
Needs or gets special therapy (e.g., physical, occupational, speech)	-	-	-	-	109 (11.0)	131 (10.2)	107 (10.7)	112 (9.4)	107 (10.7)	112 (9.4)			
Emotional, developmental, behavioral problem for which needs/gets treatment	-	-	-	-	109 (11.0)	131 (10.2)	107 (10.7)	112 (9.4)	107 (10.7)	112 (9.4)			

Note. ^aParent-report variable, percentages of children with parent-report data.

Children's Somatization Inventory-24

Children aged 8–17 years completed the CSI-24. The CSI-24 assesses the severity of nonspecific somatic symptoms (e.g., headache, chest pain, dizziness) over the past 2 weeks. The measure's internal consistency reliability was adequate among children aged 8 years and older in a prior study ($\alpha = 0.87$) (Walker et al., 2009) and in the current study ($\alpha = 0.87$).

PROMIS Short Forms

Children aged 8–17 years completed five PROMIS Pediatric short forms: Depression v.1.0 (7-item measure of sadness, loneliness, decreased positive affect and engagement), Anxiety v.1.0 (6-item measure of fear, worry, dread, and nervousness), Anger v.1.0 (9-item measure of angry mood, negative social cognitions, and efforts to control anger), Fatigue v.1.0 (10-item measure of tiredness and its impact on functioning), and Pain Interference v.1.0 (8-item measure of consequences of pain on physical, socioemotional, and cognitive functions). All measures use a 7-day recall period. The reliability and validity of these PROMIS Pediatric short forms are well established (Irwin et al., 2010, 2012b; Lai et al., 2013; Varni et al., 2010). Internal consistency reliability was adequate for all PROMIS short forms in this study (Depression: $\alpha = 0.94$; Anxiety: $\alpha = 0.91$; Anger: $\alpha = 0.92$; Fatigue: $\alpha = 0.94$; Pain Interference: $\alpha = 0.95$).

Analyses

Using Study 1 data, response frequencies, *M*s, and *SD*s were calculated for each item. Unidimensionality, local independence, and monotonicity are prerequisites for IRT modeling (Edelen & Reeve, 2007; Embretson & Reise, 2000). Item bank dimensionality was evaluated through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) 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 the first factor eigenvalue was at least four times larger than the second factor eigenvalue (Reeve et al., 2007) and in CFA by adequate single-factor model fit based on the comparative fit index ($CFI \geq 0.95$), Tucker–Lewis fit index ($TLI \geq 0.95$), and root mean square error of approximation ($RMSEA \leq 0.08$) (Hu & Bentler, 1999). 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 0 resulted in poor fit, as identified through modification indices (Bollen, 1989). When local dependence between items was observed, one of the two items was removed. Rest plots are graphs of item mean scores conditional on the total test score minus the item score. These plots were examined to confirm item monotonicity (i.e., the probability that

item endorsement increases as the measured trait increases). Nonmonotonic items were removed.

We looked for differential item functioning (DIF), meaning that people with equivalent levels of a latent variable (e.g., psychological stress experiences) respond differently to a given item 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 years), race (White vs. non-White), and ethnicity (Hispanic vs. non-Hispanic) (Carle, 2010).

Next, we used multigroup IRT to make use of Study 1 and Study 2 samples in the estimation of IRT parameters (Millsap & Yun-Tein, 2004). IRT underlies a set of generalized linear models and associated statistical procedures that connect observed item responses to the measured “ability” (e.g., level of stress experiences). When applied to the stress experiences items, IRT parameters reflect the degree to which items discriminate among children with different stress levels (*a*) and the level of stress experiences measured by each item (*b*). IRT item parameters are useful for determining whether items adequately measure the full range of an underlying variable and have optimal separation between items as indicated by minimal floor/ceiling effects, gaps in coverage (i.e., no or too few items at specific points on the continuum), and item redundancy (e.g., too many items at the same point on the continuum). Analyses were conducted in Mplus 7.2 using maximum likelihood estimation with a logit link. The stress experiences metrics were statistically identified by setting the IRT mean and variance to 0 and 1, respectively for Study 2 participants (who represent the U.S. 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 stress experiences metrics were identified using the U.S. representative sample (Study 2), IRT parameters and scores estimated from those parameters can be interpreted relative to the noninstitutionalized U.S. population, 8–17 years old (for child-report) and 5–17 years old (for parent-report). The IRT parameters are shown in Tables II and III, Supplementary Tables S1 and S2. Discrimination parameters (*a*) reflect the degree to which an item differentiates between children with varying stress levels. Location parameters (b_{1-4}) reflect the level of stress experience at which there is an equal probability of endorsing two contiguous response categories (e.g., never and almost never).

Table II. Child-Report Psychological Stress Experiences Item Statistics

Item (short form)	Item descriptive statistics				IRT parameters				
	M	SD	% floor (Never)	% ceiling (Always)	<i>a</i>	<i>b</i> ₁	<i>b</i> ₂	<i>b</i> ₃	<i>b</i> ₄
I felt stressed (SF-8, SF-4)	2.52	1.17	23.43	5.90	2.55	-0.31	0.56	1.47	2.25
I felt that my problems kept piling up (SF-8, SF-4)	2.10	1.15	39.70	4.00	3.20	0.23	0.96	1.66	2.28
I felt overwhelmed (SF-8, SF-4)	2.18	1.12	34.90	3.50	2.99	0.10	0.85	1.63	2.35
I felt unable to manage things in my life (SF-8, SF-4)	1.94	1.04	43.86	2.30	3.32	0.38	1.13	1.92	2.48
Everything bothered me (SF-8)	1.88	1.06	48.83	2.10	2.42	0.33	1.12	1.96	2.83
I felt under pressure (SF-8)	2.09	1.14	40.29	3.70	2.66	0.14	0.80	1.72	2.47
I had trouble concentrating (SF-8)	2.47	1.13	23.69	5.10	2.30	-0.18	0.62	1.67	2.46
I felt I had too much going on (SF-8)	2.53	1.16	23.11	5.80	2.37	-0.18	0.59	1.56	2.36
I had trouble making decisions	2.18	1.09	33.40	3.50	2.49	0.04	0.90	1.83	2.54
I felt concerned about what was going on in my life	2.99	1.19	33.08	5.50	2.15	0.00	0.78	1.65	2.39
Small things upset me	2.21	1.10	32.44	3.70	2.14	-0.02	0.89	1.86	2.64
I forgot things	2.52	1.11	20.65	5.70	1.99	-0.49	0.52	1.65	2.47
I felt like my thinking was slower than usual	1.83	1.01	49.95	2.10	2.53	0.51	1.26	2.15	2.77
I felt unable to remember answers, even for questions I knew the answer to	2.05	1.10	41.36	2.60	2.12	0.25	1.04	1.91	2.86
I felt so upset that I could not remember what happened or what I did	1.68	1.02	61.31	1.90	2.66	0.83	1.38	2.06	2.78
I had trouble controlling my thoughts	1.93	1.04	44.93	3.50	2.10	0.37	1.23	2.14	2.94
My thoughts went very fast	2.30	1.16	31.96	2.30	2.10	-0.03	0.77	1.72	2.56
I was slow to react to things	1.71	0.95	54.64	4.40	2.48	0.64	1.48	2.27	2.93
I felt unable to react to something that bothered me	1.87	1.04	49.25	1.60	2.44	0.49	1.23	2.04	2.83

Notes. All items begin with “In the past 4 weeks...”; response categories for all items are “Never – Rarely—Sometimes—Often – Always,” coded 1–5 such that higher values indicate more severe stress experiences.; SF-8 = eight-item short form; SF-4 = four-item short form; *a* = IRT discrimination parameter; *b*_{1–4} = IRT difficulty parameters.

We selected items for short forms (version 1.0) comprising eight items (SF-8) and four items (SF-4). Items were selected based on the following criteria: (1) discrimination parameter (*a*) indicates the item will help differentiate between children with varying stress levels; (2) item location parameters (*b*_{1–4}) cover a unique range of the stress continuum (minimal redundancy with other short-form items); (3) collectively, short-form items represented key stress experiences (e.g., arousal, agitation, pain). Short-form unidimensionality was evaluated based on CFA model fit statistics and item residual correlations. Item bank and short-form internal consistency and test–retest reliability were evaluated using Cronbach’s α and intraclass correlations, respectively. In IRT-based scores, reliability is conditional on score estimates and therefore, differs across the range of the measured concept. Marginal reliabilities equal $1 - SE^2$, where *SE* is the standard error of each estimated theta score (on the $M = 0$, $SD = 1$ metric). Marginal reliabilities were plotted by total test scores (theta scores converted to *T*-scores) to assess measurement precision across a

range of stress levels ($T = 30 - 90$) (Green, Bock, Humphres, & Linn, 1984).

We calculated full item 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) (Bock & Mislevy, 1982; Choi, 2009). We then transformed these scores into *T*-scores ($T = \Theta * 10 + 50$). A score of 50 represents the average level of stress experiences for children based on a national sample given our calibration method. Thus, scores can be interpreted relative to a U.S. population *M* of 50 and *SD* of 10. Scores can be converted to percentile ranks for the U.S. general population using *T*-distribution tables.

We calculated Pearson correlation coefficients (*r*) to assess child–parent agreement on the Psychological and Physical Stress Experiences measures. Convergent validity was evaluated by calculating Pearson correlation coefficients for scores on the stress experiences and validation measure. For Psychological Stress, we expected to observe large associations with PSS10 and PROMIS Depression, Anxiety, and Anger scales. We expected Physical Stress to have large associations

Table III. Child-Report Physical Stress Experiences Item Statistics

Item (short form)	Item descriptive statistics				IRT parameters				
	M	SD	% floor (Never)	% ceiling (Always)	a	b ₁	b ₂	b ₃	b ₄
My heart beat faster than usual, even when I was not exercising or playing hard (SF-8, SF-4)	1.41	0.82	74.39	0.92	2.48	1.14	1.88	2.72	3.35
I had trouble breathing, even when I was not exercising or playing hard (SF-8, SF-4)	1.34	0.75	78.62	0.98	2.40	1.38	2.06	2.98	3.45
My body shook (SF-8, SF-4)	1.38	0.79	75.64	1.03	2.25	1.36	2.12	2.87	3.60
I had pain that really bothered me (SF-8, SF-4)	1.70	1.05	61.04	2.77	1.82	0.69	1.50	2.40	3.18
My muscles felt tight (SF-8)	1.83	1.05	51.98	2.55	1.71	0.50	1.43	2.52	3.40
My mouth was dry (SF-8)	1.76	0.97	52.63	1.36	1.68	0.54	1.57	2.70	3.83
I had a headache (SF-8)	2.07	1.10	40.21	3.20	1.34	-0.12	1.03	2.51	3.73
My back hurt (SF-8)	1.94	1.16	49.76	4.40	1.63	0.43	1.31	2.34	3.10
I felt as if I needed to move my legs a lot	1.76	1.09	59.14	2.82	1.58	0.71	1.57	2.43	3.41
A sudden noise made me jump	1.83	1.08	53.12	3.04	1.46	0.45	1.47	2.58	3.50
I could not stay still for long	2.12	1.25	44.28	6.19	1.13	0.06	1.15	2.23	3.32
My breathing was fast, even when I was not exercising or playing hard	1.39	0.81	76.07	0.92	2.75	1.28	1.96	2.69	3.39
My palms were sweaty	1.80	1.05	54.04	2.60	1.47	0.49	1.54	2.64	3.59
I was sweaty, even when I was not exercising or playing hard	1.47	0.89	71.84	1.63	1.85	1.20	2.02	2.84	3.55
My heart pounded, even when I was not exercising or playing hard	1.63	0.96	62.18	1.52	2.11	0.78	1.64	2.55	3.41
I felt dizzy	1.54	0.91	67.93	1.25	2.26	1.00	1.75	2.61	3.44
My breathing was fast	1.44	0.83	72.60	1.19	2.43	1.18	1.91	2.89	3.39
My hands shook	1.40	0.83	75.47	1.68	2.37	1.30	2.09	2.78	3.24
My legs shook	1.36	0.80	78.30	1.52	2.52	1.40	2.09	2.81	3.25
My neck hurt	1.57	0.94	65.87	1.57	1.97	0.94	1.77	2.67	3.47
I had a bad stomach ache	1.74	0.99	55.51	1.47	1.40	0.54	1.66	2.85	4.17
My appetite changed	1.63	1.00	64.46	1.84	1.79	0.88	1.63	2.61	3.54
I felt some food coming up into my throat	1.41	0.79	73.36	1.09	1.77	1.28	2.21	3.26	3.89
I threw up	1.30	0.68	80.03	0.60	1.49	1.68	2.65	3.80	4.67
I had trouble staying asleep	1.85	1.12	53.93	3.74	1.33	0.48	1.48	2.58	3.48
My neck felt tight	1.62	0.98	64.46	1.79	1.85	0.89	1.68	2.68	3.48

Notes. All items begin with “In the past 4 weeks...”; response categories for all items are “Never – Rarely – Sometimes – Often – Always,” coded 1–5 such that higher values indicate more severe stress experiences.; SF-8 = eight-item short form; SF-4 = four-item short form; a = IRT discrimination parameter; b₁₋₄ = IRT difficulty parameters.

with CSI-24 and PROMIS Fatigue and Pain Interference scales. Construct validity was evaluated based on associations between children’s special health-care need status and stress experience scores. We conducted stepwise regression analyses in which child and family sociodemographic covariates (age, gender, race, ethnicity, family income, parental education) were entered in the first block and child special health-care needs (medication use, service use, functional limitations, therapy, emotional/developmental/behavioral problems) were entered in the second block. Dependent variables were child- and parent-proxy Psychological and Physical Stress Experiences short-form scores. We hypothesized that children with special health-care needs would have higher stress experience scores. We also expected child age to be positively correlated with psychological stress.

Results

Rates of missing data were ≤ 1% for all items. In general, the items were positively skewed, but all response categories were endorsed for every item. Descriptive statistics are shown for the final child-report items in Tables II and III (and for parent-report items in the Online Supplementary Material).

**IRT Assumptions
Psychological Stress Experiences**

A single-factor CFA model containing all 63 of the child-report psychological stress items fit the data poorly (CFI = 0.83, TLI = 0.83; RMSEA = 0.09). An EFA model retaining four orthogonal factors best met our model selection criteria and accounted for 54.5% of the variance in the items. In this model, 60 items contributed prominently to one of the four factors. Two items with nonsalient factor loadings (“I had

trouble thinking clearly” and “I had trouble paying attention”) and one item that contributed to two of the four factors (“I felt grouchy”) were removed. Of the 60 remaining items, 24 contributed to the first factor in the EFA model (eigenvalue = 30.3). These items largely reflected the perceived controllability/manageability and cognitive-perceptual disruption facets of psychological stress. The second factor (eigenvalue = 3.0) was composed of 16 items that characterize severe fear reactions (e.g., “I had to watch out for danger all the time”; “I felt terrified.”). The third factor (eigenvalue = 2.6) was composed of 10 items that describe expressions of anger or rage (“I felt really mad”; “I wanted to hurt someone”). Finally, 10 items loaded on the fourth factor (eigenvalue = 1.4), which characterized positive experiences that children with high levels of stress were expected to endorse at low levels (e.g., “I felt in control of my life”; “I felt safe”).

We removed items that contributed to the second and third factors because they are conceptually redundant with the existing PROMIS Anxiety and Anger item banks (Irwin et al., 2010, 2012b). Items with positive valence (Factor 4) such as feeling relaxed, in control, and having enough time to complete required tasks were also removed to support development of a single unidimensional psychological stress item bank. Remaining analyses focused on Factor 1 items. A single-factor CFA model composed of the 24 Factor 1 items failed to adequately fit the data. Two items had factor loadings < 0.70: “I felt distressed” (loading = 0.66) and “I felt I had too many things to do” (loading = 0.66). Modification indices revealed local dependencies between three item pairs: (1) “I felt annoyed” and “Everything bothered me”; (2) “I felt distracted” and “I had trouble concentrating”; and (3) “I forgot things I needed to remember” and “I forgot things.” Items with low loadings and the first item from each locally dependent pair were removed. A single-factor model composed of the 19 parent-proxy items that paralleled those retained in the final child-report item bank was a poor fit for the data, and EFA revealed that two factors accounted for 59.3% of the variance in these items. The majority of items (12 of 19) contributed to the first factor (eigenvalue = 11.3), which represented perceptions that demands are uncontrollable, unpredictable, and overloading. These items converged on a unidimensional psychological stress construct (CFI = 0.96, TLI = 0.95, RMSEA = 0.11). The seven remaining items, which contributed to Factor 2 (eigenvalue = 1.4), measured children’s cognitive-perceptual stress reactions (e.g., “My child was slow to react to things”; “My child forgot things.”). These items were removed from the final parent-proxy item bank. Rest plots showed that no

child- and parent-report psychological stress items violated the monotonicity assumption.

Physical Stress Experiences

A single-factor CFA model including all 40 child-report physical stress items fit the data poorly (CFI = 0.81, TLI = 0.80; RMSEA = 0.11). In EFA, a two-factor solution best met our model fit criteria and accounted for 59.9% of the variance in items. The first factor had an eigenvalue of 22.8. Eleven items that contributed to Factor 2 (eigenvalue = 2.9) were eliminated to support development of a single unidimensional physical stress item bank. Most Factor 2 items characterized sleep disturbance (e.g., trouble falling asleep; nightmares). Factor 1 items assessed the physical stress experiences of agitation, arousal, tension, pain, and gastrointestinal distress. It included a single item about sleep: “trouble staying asleep.” In a single-factor CFA model including the 29 Factor 1 items, one item (“I had to go to the bathroom more than usual”) had a low factor loading (0.52), and modification indices revealed three locally dependent items: “It was hard for me to stay in my seat at school”; “It was hard for me to sit still”; and “I could not stay still for long.” We removed the poor fitting item and two of the three locally dependent items, retaining “I could not stay still for long” because it describes agitation in the most general and widely applicable terms. A single-factor CFA model adequately fits the data for the remaining 26 child-report physical stress items (CFI = 0.95, TLI = 0.94, RMSEA = 0.09). Likewise, a single-factor model composed of the 26 parent-proxy items that paralleled those retained in the final child-report item bank adequately fit the data (CFI = 0.95, TLI = 0.95, RMSEA = 0.07), and no local dependency was observed among items. Rest plots showed that no child- and parent-report physical stress items violated the monotonicity assumption.

Differential Item Functioning

For both psychological and physical stress, we observed statistically significant DIF in multiple child- and parent-proxy item thresholds, but no differences in discrimination parameters as a function of child race, ethnicity, age, and gender. Nonetheless, DIF did not substantively impact total test scores. For all grouping variables and both reporters, full bank, SF-8, and SF-4 scores ignoring and adjusting for DIF were nearly perfectly correlated (> 0.99) (see [Supplementary Material](#)). DIF impact was positively associated with stress levels, but even at the 99th percentile, the mean change in scores after adjusting for DIF was < 0.20 SD units (or two points on the *T*-score metric) for all measures. As a result, no items were eliminated because of DIF.

Table IV. Child-Report Psychological and Physical Stress Experiences Full Bank and Short-Form Descriptive Statistics, CFA Model Fit, Reliability and Concurrent Validity

Analysis type	Psychological stress					Physical stress				
	Study 1 (n)	Study 2 (n)	Full bank	SF-8	SF-4	Study 1 (n)	Study 2 (n)	Full bank	SF-8	SF-4
Descriptive statistics (Study 2)										
M	–	994	–	49.99	50.10	–	1,004	–	50.01	50.86
SD	–	994	–	9.33	9.06	–	1,004	–	8.59	7.76
Minimum	–	994	–	37.00	39.47	–	1,004	–	39.41	44.46
Maximum	–	994	–	78.60	78.43	–	1,004	–	75.34	80.73
Skew	–	994	–	0.22	9.06	–	1,004	–	0.44	1.03
Descriptive statistics (Study 1)										
M	1,874	–	54.64	54.48	54.13	1,843	–	52.81	52.73	52.10
SD	1,874	–	8.92	8.92	8.88	1,843	–	11.17	10.26	9.18
Minimum	1,874	–	34.21	37.00	39.47	1,843	–	35.16	39.41	44.46
Maximum	1,874	–	86.13	81.83	78.43	1,843	–	89.56	87.34	84.85
Skew	1,874	–	–0.18	–0.03	0.00	1,843	–	0.16	0.37	0.91
Confirmatory factor model fit										
CFI	1,874	–	0.95	0.99	0.99	1,843	–	0.95	0.98	0.99
TLI	1,874	–	0.94	0.99	0.99	1,843	–	0.94	0.97	0.99
RMSEA	1,874	–	0.10	0.09	0.11	1,843	–	0.09	0.10	0.09
Reliability										
Internal consistency (α)	1,874	–	0.95	0.92	0.87	1,843	–	0.95	0.87	0.81
Test–retest reliability (ICC)	104	–	0.70	0.71	0.72	124	–	0.70	0.68	0.70
Agreement with parent-report (r)										
Parent-report full bank	821	–	0.62	0.61	0.58	820	–	0.63	0.59	0.55
Parent-report SF-8	821	–	0.61	0.61	0.58	820	–	0.59	0.60	0.55
Parent-report SF-4	821	–	0.59	0.59	0.57	820	–	0.55	0.54	0.57
Concurrent validation (r)										
Perceived stress ^a	171	648	0.79	0.80	0.76	170	–	0.63	0.59	0.64
Somatization ^b	–	–	–	–	–	–	468	–	0.69	0.57
Depression ^c	211	–	0.75	0.71	0.67	211	–	0.60	0.66	0.72
Anxiety ^c	214	–	0.72	0.68	0.65	214	–	0.52	0.58	0.67
Anger ^c	212	–	0.72	0.68	0.62	212	–	0.57	0.63	0.66
Fatigue ^c	209	–	0.76	0.74	0.70	209	–	0.72	0.67	0.60
Pain interference ^c	212	–	0.59	0.52	0.57	212	–	0.72	0.69	0.65

Notes. CFI = comparative fit index; r = Pearson’s correlation; RMSEA = root mean square error of approximation; TLI = Tucker–Lewis Index; α = Cronbach’s alpha; ICC = intraclass correlation coefficient; ^aPerceived Stress Scale (PSS10); ^bChildren’s Somatization Inventory-24 (CSI-24); ^cPROMIS pediatric short forms.

In Study 1, unique randomly selected groups of children completed the PSS10, PROMIS Depression/Anxiety/Anger short forms, or PROMIS Fatigue/Pain interference short forms. In Study 2, unique randomly selected subgroups of children completed PSS10 or CSI-24.

Item Bank Calibration

Discrimination and location parameters for the graded response models are shown in Tables II and III, [Supplementary Tables S1 and S2](#). All items contributed to the assessment of the severity of psychological or physical stress experiences. For a single parent-proxy item (“My child’s breathing was fast”), the “Often” and “Always” response categories measured the same severity level of physical stress. Therefore, we collapsed these response categories and assigned the same numeric value for this item. Item location parameters reflected a sizable range of psychological (child: $\Theta = -0.49$ to 2.93; parent: $\Theta = -0.36$ to 3.55) and physical stress experiences (child: $\Theta = -0.12$ to 4.67; parent: $\Theta = 0.28$ –4.84).

Short Forms

Items that maximized the conceptual coverage and range of precisely measured stress experiences were

selected for eight- and four-item short forms (version 1.0) (Tables II and III, [Supplementary Tables S1 and S2](#)). CFA fit statistics supported short-form unidimensionality, and all child- and parent-report instruments were found to have adequate or good internal consistency and test–retest reliability ([Table IV](#), [Supplementary Table S3](#)). Short-form scoring procedures are included in the [Online Supplementary Material](#).

Precision

Marginal reliability plots demonstrated that the full bank and short forms have acceptable levels of precision across a range of stress experiences ([Figures 1 and 2](#), [Supplementary Figures S1 and S2](#)). Instrument versions with more items provided greater precision than those with fewer items. As is common of symptom measures, all versions were better at discriminating

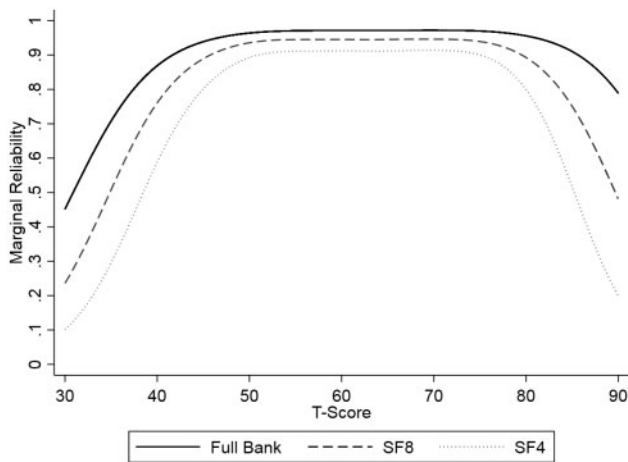


Figure 1. Child-report Psychological Stress Experiences Marginal Reliability.

among children with above-average ($T > 0.50$) stress experiences.

Child and Parent Agreement

Pearson correlations revealed moderate agreement between child- and parent-report versions of the same psychological or physical stress measures (Table IV). Agreement was comparable for the full bank and SF-8 versions of the measures and slightly poorer for the SF-4s.

Validity

Correlations between psychological/physical stress and validation measures are shown in Table IV and Supplementary Table S3. Large associations ($r > .7$) were observed for scores on psychological stress, and all validation measures except PROMIS Pain Interference. Physical stress scores had large associations with PROMIS Fatigue and Pain Interference scores, and moderate associations ($0.5 > r < .7$) with scores on all other validation measures. Adolescents (aged 13–17 years) experienced more child- and parent-reported psychological stress than children aged 8–12 (by child- and parent-report). Parents of younger children (aged 5–7 years) reported that their children experienced less psychological stress. After controlling for child age, gender, race, ethnicity, family income, and parental education attainment, children with four of five special health-care qualifying criteria had higher stress levels (Table V, Supplementary Table S4). Prescription medication need or use predicted increased psychological stress by child-report and physical stress by both child- and parent-report. Need or use of medical care, mental health, or educational services was associated with higher psychological stress according to both reporters. Functional limitations and emotional/

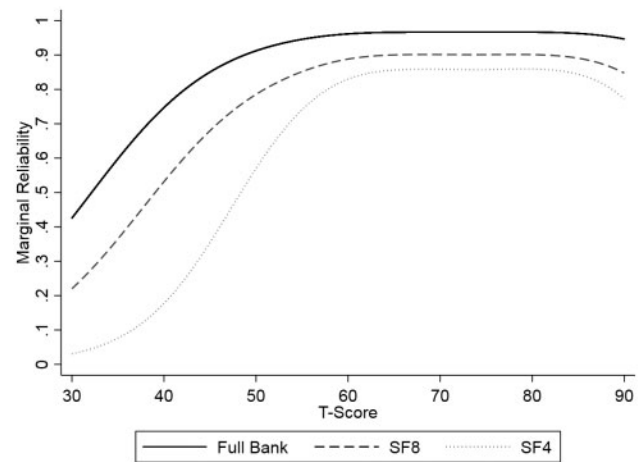


Figure 2. Child-report Physical Stress Experiences Marginal Reliability.

developmental/behavioral problems predicted greater child- and parent-proxy-reported Psychological and Physical Stress Experiences.

Discussion

Developmentally sensitive, precise, and efficient measures of children's stress experiences are needed to support early detection of atypical stress and to assess intervention effectiveness. The PROMIS Pediatric Psychological and Physical Stress Experiences instruments were developed to meet this need for clinical research and practice. The item banks and short forms (version 1.0) were created using a rigorous mixed-method approach consistent with multiple patient-reported outcome measure development standards (Acaster, Cimms, & Lloyd, 2012; Mokkink et al., 2010; National Quality Forum, 2013; Patrick et al., 2007). As previously described, the instruments are theoretically grounded and consistent with child, parent, and clinician perspectives on perceived stress (Bevans et al., 2013). Data collected from over 5,000 children and parents were used to evaluate the psychometric properties of the items and to calibrate them using IRT.

The final child-report psychological stress experiences instruments assess stress-related disruptions to cognitive processes, feeling overwhelmed, and lack of perceived control. Factor analyses demonstrated that these experiences can be distinguished from nervousness, worry, and irritability. The Psychological Stress Experiences instruments complement existing PROMIS pediatric Anxiety, Depression, and Anger measures, and thereby expand the measurement system's coverage of stress-related mental health outcomes that are meaningful to children and families. Moreover, the PROMIS child-report psychological

Table V. Summary of Stepwise Regression for Variables Predicting Child-Report Psychological and Physical Stress Experiences

Participant characteristic	Psychological stress						Physical stress					
	SF-8			SF-4			SF-8			SF-4		
	R ²	ΔR ²	β	R ²	ΔR ²	β	R ²	ΔR ²	β	R ²	ΔR ²	β
Step 1: Sociodemographics	0.05	0.05		0.05	0.05		0.01	0.01		0.01	0.01	
Age (years) (ref: 8–12)												
13–17			2.62***			2.89***			0.90			0.00
Gender (ref: boys)												
Girls			2.34***			1.85**			0.55			0.63
Race (ref: White)												
African-American			−1.76*			−2.53**			0.47			1.03
Asian			0.03			−0.35			3.00*			3.05*
Other			0.60			0.66			−0.47			−0.20
Ethnicity (ref: non-Hispanic)												
Hispanic			−0.22			−0.36			0.84			1.20
Annual family income (ref: ≥\$40K)												
<\$40,000			1.60*			1.79**			−0.03			−0.23
Parent educational attainment (ref: College degree or higher)												
High School/GED or less			−0.84			−0.72			−0.72			−0.34
Some college/associate’s degree			0.52			0.10			−0.24			0.33
Step 2: Special healthcare need (ref: does not have SHCN)	0.15	0.10		0.16	0.11		0.10	0.09		0.12	0.11	
Needs/uses medication (prescribed)			1.98*			1.70*			4.23***			4.05***
Needs/uses medical, mental health, or psychological services			3.39**			3.97**			−0.73			0.64
Has functional limitation			4.60**			4.57**			4.50**			4.51**
Needs or gets special therapy (e.g., physical, occupational, speech)			−2.51			−2.74			−1.79			−2.09
Emotional, developmental, behavioral problem for which needs/gets treatment			4.43***			4.29***			3.16**			2.60**

Notes. ref = referent group; for entire model psychological SF-8: $F(14,961) = 11.9***$; SF-4: $F(14,943) = 12.8***$; physical SF-8: $F(14,955) = 7.7***$; SF-4: $F(14,933) = 9.4***$; * $p < .05$; ** $p < .01$; *** $p < .001$.

stress instruments extend the perceived stress concept as measured by the PSS10 to include stress-related cognitive–perceptual disruptions (e.g., trouble concentrating, difficulty controlling thoughts). Whereas cognitive–perceptual disruption items contributed to the measurement of a single psychological stress construct by child-report, some of these items comprised a unique dimension by parent-proxy report. As a result, the final parent-proxy psychological stress item bank contains fewer cognitive–perceptual disruption items than the child-report version. Parents or other external observers may underestimate cognitive–perceptual consequences of stress such as memory loss or changes in processing speed. Indeed, parents endorsed cognitive–perceptual disruption items at lower levels than children.

The physical stress experiences instruments assess arousal (e.g., sensory alertness, muscle potential), agitation (e.g., restlessness, fidgetiness), pain, gastrointestinal distress, and difficulty staying asleep. Users

should interpret physical stress scores with awareness of a child’s overall health status, as some items reflect symptoms of chronic conditions. Indeed, children who need or use prescription medication (possibly to treat a chronic condition) experience more physical stress. However, the unidimensionality of physical stress experiences stemming from cardiovascular, respiratory, musculoskeletal, and gastrointestinal systems suggests that the physical stress items measure bodily manifestations of stress (rather than disease-specific symptoms).

As expected, children with special health-care needs had more severe stress experiences. Stress was most pronounced for children who needed or used medical, mental health, or behavioral services and those with functional limitations and/or emotional, developmental, or behavioral problems. These associations are consistent with prior research demonstrating that stress predisposes children to physical and mental disorders (Barr, Boyce, & Zeltzer, 1996;

Graham et al., 1986; Rutter, 1989). Conversely, treatment regimens and other consequences of having a medical, developmental, or behavioral condition are significant stressors that increase children's risk for Psychological and Physical Stress Experiences (Reigada et al., 2011). Despite these associations, many children with special health-care needs reported moderate or below-average levels of stress. The PROMIS pediatric stress experience measures could be used to identify children who need support managing events that exceed their adaptive capacity. Moreover, because the measures assess a broad range of stress experiences (including mild and moderate stress reactions), they will be useful for targeting and evaluating interventions that reduce or prevent symptoms of stress-related conditions (e.g., anxiety, depression).

Limitations and Recommendations for Future Work

Both the Psychological and Physical Stress Experiences instruments are more precise measures of above-average levels of stress. The addition of items that capture less severe stress response may strengthen the measures' sensitivity to change in stress experiences resulting from effective stress-reduction interventions. Longitudinal data are needed to evaluate the instruments' predictive validity and responsiveness to change. These data could also be used to test the hypothesis that stress experiences reflect early manifestations of more severe health concerns. Although we purposively sampled to maximize representation of sociodemographically diverse families and children with a range of health conditions, the content and clinical validity of the measures should be further evaluated in other populations and contexts. We selected items for version 1.0 of eight- and four-item short forms to maximize the measures' precision across a wide range of stress levels (least to most severe). However, alternative short forms composed of a different set of items from the calibrated item banks may be preferred for children with a particular attribute (e.g., clinical condition) or given the purpose of the assessment (e.g., to screen for high stress levels). Alternative short-form versions that include different items from the same IRT-calibrated item bank can be scored on the same metric, and therefore compared (Broderick et al., 2013). In the future, the process for generating alternative short-form versions would be strengthened by evaluating the degree to which experts agree on the selection of specific items (e.g., interrater reliability). Finally, although all questionnaires were self-administered, the studies collected data in different locations (home and school) and used different methods (computerized vs. paper-and-pencil). The degree to which assessment modality impacted the instruments' measurement properties is unknown.

Conclusion

The PROMIS Pediatric Psychological and Physical Stress Experiences item banks and short forms (version 1.0) provide precise measurement across a wide range of stress severity levels irrespective of child age, gender, race/ethnicity, and chronic disease status. Known-group comparisons and convergence with existing measures of perceived stress, emotional distress, fatigue, and pain provided evidence of construct validity. The scores are interpretable relative to the U.S. population.

Supplementary Data

Supplementary data can be found at: <http://www.jpepsy.oxfordjournals.org/>.

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 Tulskey, 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.healthmeasures.org for additional information on the PROMIS® initiative.

Conflicts of interest: None declared.

References

- Acaster, S., Cimms, T., & Lloyd, A. (2012). *Development of a methodological standards report #3: The design and selection of patient-reported outcomes measures (PROMs) for use in patient centered outcomes research*. Retrieved from <http://www.pcori.org/assets/The-Design-and-Selection-of-Patient-Reported-Outcomes-Measures-for-Use-in-Patient-Centered-Outcomes-Research1.pdf>
- Barr, R. G., Boyce, W. T., & Zeltzer, L. K. (1996). The stress-illness association in children: A perspective from the biobehavioral interface. In R. J. Haggerty, L. R. Sherrold, N. Garnezy, & M. Rutter (Eds.), *Stress, risk, and resilience in children and adolescents: Processes, mechanisms, and interventions*. Cambridge, UK: Cambridge University Press.
- Bethell, C., Read, D., Stein, R., Blumberg, S., Wells, N., & Newacheck, P. (2002). Identifying children with special health care needs: Development and evaluation of a short screening instrument. *Ambulatory Pediatrics*, 2, 38–48.
- Bevans, K., Gardner, W., Pajer, K., Riley, A., & Forrest, C. (2013). Qualitative development of the PROMIS(R) pediatric stress response item banks. *Journal of Pediatric Psychology*, 38, 173–191.
- Bevans, K. B., Moon, J., Riley, A. W., & Forrest, C. B. (2010). Conceptual and methodological advances in child reported outcomes measurement. *Pharmacoeconomics and Outcomes Research*, 10, 385–396.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431–444.
- Bollen, K. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Boyce, W. T., Chesney, M., Alkon, A., Tschann, J. M., Adams, S., Chesterman, B., ... Wara, D. (1995). Psychobiologic reactivity to stress and childhood respiratory illness: Results of two prospective studies. *Psychosomatic Medicine*, 57, 411–422.
- Broderick, J. E., Morgan-DeWitt, E., Rothrock, N., Crane, P., & Forrest, C. B. (2013). Advances in patient reported outcomes: The NIH PROMIS measures. *eGEMs (Generating Evidence and Methods to Improve Patient Outcomes)*, 1, 1015. Retrieved from <http://doi.org/10.13063/2327-9214.1015>
- Burke, C. A. (2010). Mindfulness-based approaches with children and adolescents: A preliminary review of current research in an emergent field. *Journal of Child and Family Studies*, 19, 133–144. Retrieved from <http://doi.org/10.1007/s10826-009-9282-x>
- Carle, A. (2010). Mitigating systematic measurement error in comparative effectiveness research in heterogeneous populations. *Medical Care*, 48, S68.
- Choi, S. (2009). Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. *Applied Psychological Measurement*, 33, 644–645.
- Cohen, S., Kessler, R. C., & Gordon, L. U. (1997). *Measuring stress: A guide for health and social scientists*. New York, NY: Oxford University Press.
- Cohen, S., & Williamson, G. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan & S. Oskamp (Eds.), *The social psychology of health* (pp. 31–67). Newbury Park, CA: Sage.
- Dennis, J. (2010). *KnowledgePanel®: Processes & procedures contributing to sample representativeness & tests for self-selection bias*. Retrieved from <http://www.knowledgenetworks.com/ganp/docs/KnowledgePanelR-Statistical-Methods-Note.pdf> Retrieved June 6, 2016.
- 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, 4174–4183. Alexandria, VA: American Statistical Association.
- Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16, 5–18.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Franke, H. (2014). Toxic stress: Effects, prevention and treatment. *Children*, 1, 390–402. Retrieved from <http://doi.org/10.3390/children1030390>
- Garnezy, N. (1986). Developmental aspects of children's responses to the stress of separation and loss. In M. Rutter, C. E. Izard, & P. B. Reid (Eds.), *Depression in young people: Developmental and clinical perspectives*. New York, NY: Guilford Press.
- Gould, L. F., Hussong, A. M., & Keeley, M. L. (2008). The adolescent coping process interview: Measuring temporal and affective components of adolescent responses to peer stress. *Journal of Adolescence*, 31, 641–657.
- Graham, N., Douglas, R. M., & Ryan, P. (1986). Stress and acute respiratory infection. *American Journal of Epidemiology*, 124, 389–401.
- Green, B. F., Bock, D., Humphres, R. L., & Linn, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21, 347–360.
- Harnett, P. H., & Dawe, S. (2012). The contribution of mindfulness-based therapies for children and families and proposed conceptual integration. *Child and Adolescent Mental Health*, 17, 195–208. Retrieved from <http://doi.org/10.1111/j.1475-3588.2011.00643.x>
- 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.
- Irwin, D. E., Gross, H. E., Stucky, B. D., Thissen, D., Morgan Dewitt, E., & Lai, J. S. (2012a). Development of six PROMIS pediatrics proxy-report item banks. *Health Qual Life Outcomes*, 10, 10–22.
- Irwin, D. E., Stucky, B., Langer, M. M., Thissen, D., DeWitt, E. M., Lai, J. S., ... DeWalt, D. A. (2010). An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Quality of Life Research*, 19, 595–607.

- Irwin, D. E., Stucky, B. D., Langer, M. M., Thissen, D., DeWitt, E. M., Lai, J. S., ... DeWalt, D. A. (2012b). PROMIS pediatric anger scale: An item response theory analysis. *Quality of Life Research*, *21*, 697–706.
- Klare, G. (1976). A second look at the validity of readability formulas. *Journal of Reading Behavior*, *8*, 129–152.
- Kugler, B. B., Bloom, M., Kaercher, L. B., Truax, T. V., & Storch, E. A. (2012). Somatic symptoms in traumatized children and adolescents. *Child Psychiatry and Human Development*, *43*, 661–673. Retrieved from <http://doi.org/10.1007/s10578-012-0289-y>
- Kupst, M. J., Butt, Z., Stoney, C. M., Griffith, J. W., Salsman, J. M., Folkman, S., & Cella, D. (2015). Assessment of stress and self-efficacy for the NIH toolbox for neurological and behavioral function. *Anxiety, Stress, and Coping*, *28*, 531–544. Retrieved from <http://doi.org/10.1080/10615806.2014.994204>
- Lai, J. S., Stucky, B. D., Thissen, D., Varni, J. W., DeWitt, E. M., Irwin, D. E., ... DeWalt, D. A. (2013). Development and psychometric properties of the PROMIS® pediatric fatigue item banks. *Quality of Life Research*, *22*, 2417–2427. Retrieved from <http://doi.org/10.1007/s11136-013-0357-1>
- Lasch, K. E., Marquis, P., Vigneux, M., Abetz, L., Arnould, B., Bayliss, M., ... Rosa, K. (2010). PRO development: Rigorous qualitative research as the crucial foundation. *Quality of Life Research*, *19*, 1087–1096. Retrieved from <http://doi.org/10.1007/s11136-010-9677-6>
- Laurent, J., Catanzaro, S. J., & Joiner, T. E. (2004). Development and preliminary validation of the physiological hyperarousal scale for children. *Psychological Assessment*, *16*, 373–380. Retrieved from <http://doi.org/10.1037/1040-3590.16.4.373>
- Lohr, S. (2009). *Sampling: Design and analysis*. Toronto, ON: Nelson Education.
- Magasi, S., Ryan, G., & Revicki, D. (2012). Content validity of patient-reported outcome measures: Perspectives from a PROMIS meeting. *Quality of Life Research*, *21*, 739–736.
- March, J. S., Parker, J. D., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The multidimensional anxiety scale for children (MASC): Factor structure, reliability, and validity. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 554–565.
- Matza, L. S., Patrick, D. L., Riley, A. W., Alexander, J. J., Rajmil, L., Pleil, A. M., & Bullinger, M. (2013). Pediatric patient-reported outcome instruments for research to support medical product labeling: Report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. *Value in Health*, *16*, 461–479. Retrieved from <http://doi.org/10.1016/j.jval.2013.04.004>
- Millsap, R., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*, 479–513.
- Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., ... de Vet, H. C. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*, *10*, 22. Retrieved from <http://doi.org/10.1186/1471-2288-10-22>
- National Quality Forum (2013). *Patient Reported Outcomes (PROs) in performance measurement*. Retrieved from http://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx
- Patrick, D. L., Burke, L. B., Powers, J. H., Scott, J. A., Rock, E. P., Dawisha, S., ... Kennedy, D. L. (2007). Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value in Health*, *10*, S125–S137. Retrieved from <http://doi.org/10.1111/j.1524-4733.2007.00275.x>
- Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., ... Cella, D.; PROMIS Cooperative Group (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*, S22–S31.
- Reigada, L. C., Bruzzese, J., Benkov, K. J., Levy, J., Waxman, A. R., Petkova, E., & Warner, C. M. (2011). Illness-specific anxiety: Implications for functioning and utilization of medical services in adolescents with inflammatory. *Journal for Specialists in Pediatric Nursing*, *16*, 207–215. Retrieved from <http://doi.org/10.1111/j.1744-6155.2011.00292.x>
- Reynolds, C. R., & Paget, K. D. (1981). Factor analysis of the revised children's manifest anxiety scale for blacks, whites, males, and females with a national normative sample. *Journal of Consulting and Clinical Psychology*, *49*, 352–359. Retrieved from <http://doi.org/10.1037/0022-006X.49.3.352>
- Rutter, M. (1989). Pathways from childhood to adult life. *Journal of Child Psychology and Psychiatry*, *30*, 23–51.
- Shonkoff, J. P., Garner, A. S., Siegel, B. S., Dobbins, M. I., Earls, M. F., Garner, A. S., ... Wegner, L. M. (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics*, *129*, e232–e246. Retrieved from <http://doi.org/10.1542/peds.2011-2663>
- Storch, E. A., Merlo, L. J., Keeley, M. L., Grabill, K., Milsom, V. A., Geffken, G. R., & Goodman, W. K. (2008). Somatic symptoms in children and adolescents with obsessive-compulsive disorder: Associations with clinical characteristics and cognitive-behavioral therapy response. *Behavioural and Cognitive Psychotherapy*, *36*, 283–297.
- Varni, J. W., Stucky, B. D., Thissen, D., Dewitt, E. M., Irwin, D. E., Lai, J. S., ... Dewalt, D. A. (2010). PROMIS pediatric pain interference scale: An item response theory analysis of the pediatric pain item bank. *The Journal of Pain*, *11*, 1109–1119. Retrieved from <http://doi.org/10.1016/j.jpain.2010.02.005>. PROMIS
- Walker, L., & Garber, J. (1992). *Children's somatization inventory: Preliminary manual*. Nashville, TN: Vanderbilt University Medical Center.
- Walker, L. S., Beck, J. E., Garber, J., & Lambert, W. (2009). Children's somatization inventory: psychometric properties of the revised form (CSI-24). *Journal of Pediatric Psychology*, *34*, 430–440.
- Watt, M. J., Weber, M. A., Davies, S. R., & Forster, G. L. (2017). Impact of juvenile chronic stress on adult cortico-accumbal function: Implications for cognition and addiction. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *79*, 136–154. Retrieved from <http://doi.org/10.1016/j.pnpbp.2017.06.015>