Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Anxiety Stress Coping. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Anxiety Stress Coping. 2015 September ; 28(5): 531–544. doi:10.1080/10615806.2014.994204.

Assessing Stress and Self-Efficacy for the NIH Toolbox for Neurological and Behavioral Function

Mary Jo Kupst,
Medical College of Wisconsin, Department of Pediatrics

Zeeshan Butt,
Department of Medical Social Sciences, Comprehensive Transplant Center and Institute for Healthcare Studies, Northwestern University

Catherine M. Stoney,
National Institutes of Health

James W. Griffith,
Department of Medical Social Sciences, Northwestern University

John M. Salsman,
Department of Medical Social Sciences, Robert H. Lurie Comprehensive Cancer Center of Northwestern University

Susan Folkman, and
Department of Medicine, University of California San Francisco (Emeritus)

David Cella
Department of Medical Social Sciences, Northwestern University

Many measures of neurological and behavioral health are available to assess changes in health and function over time. Typically, however, such measures are not incorporated into large-scale epidemiological studies and clinical trials because they are expensive, lengthy, limited to certain age groups or populations, or not easily administered (Gershon et al., 2013). Including neurological and behavioral function measures in these large and expensive epidemiological studies could improve understanding of how neurological and behavioral health factors are related to diverse outcomes and do so cost-effectively and with minimal additional participant burden.

The NIH Toolbox for the Assessment of Neurological and Behavioral Function (NIH Toolbox; nihtoolbox.org) initiative was developed to identify, create when necessary, and validate measures in the broad domains of cognitive function, emotional health, motor function, and sensory function for use large-scale epidemiological studies (Gershon, et al., 2010; Gershon, et al. 2013; Salsman, et al., 2013). Based on a multi-level process, involving content experts and potential end-users (Nowinski, Victorson, Debb, & Gershon, 2013), several subdomains were identified within each of these broad domains. For example, the

Corresponding Author: Mary Jo Kupst, Ph.D., Emerita Professor of Pediatrics, Medical College of Wisconsin, Pediatric Hematology/Oncology, 8701 Watertown Plank Road, MFRC 3018, Milwaukee WI 53226. 815-347-8648. FAX: 414-955-6543. mkupst@mcw.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

emotional health domain was comprised of four subdomains: Negative Affect, Positive Affect, Social Relationships, and Stress and Self-Efficacy subdomains, the latter of which is the focus of this paper. In this article we present the process through which Stress and Self-Efficacy candidate measures were identified, evaluated, and selected for further norming and psychometric testing.

Conceptual Framework

Stress

The relevance of stress to disease and health has been demonstrated in both human and animal models (Cohen, Janicki-Deverts, & Miller, 2007; Segerstrom and Miller, 2004). For example, stress can hamper adherence to medication and other medical recommendations, and it can lead to increases in health-damaging behavior. Among individuals who are well, stress can have a direct impact on health through its effects on physiological and behavioral functioning. It is related to risk of developing cardiovascular diseases, immune dysfunction, and certain cancers, through increased autonomic nervous system activity, poor vagal tone, chronically elevated blood cortisol levels, and decreased immunocompetence (Antoni et al., 2006; Cohen & Williamson, 1988; Raposa, Hammen, Brennan, O'Callaghan, & Najman, 2014; Steptoe & Kivimaki, 2012).

Stress involves both environmental (events or experiences that are assumed to be stressors) and subjective factors (an individual's perception of stress). For the purposes of this project, where developed measures may later be used to assess many diverse types of stressor, stress refers to the experience of psychosocial stress, or perceived stress. The subjective experience of psychological stress is influenced by the individual's appraisal of the personal significance of a given event in relation to his or her values and goals and options and resources for controlling the outcome (Folkman, 1984). Psychological stress is experienced when an individual perceives environmental or internal demands that are personally threatening and exceed his or her perceived adaptive capacity (Cohen, Kessler, & Gordon, 1997; Folkman, 2010; Lazarus & Folkman, 1984). Rather than focusing primarily on specific stressors, this definition focuses on a person's perception of events or experiences that take into account both environmental factors and one's perceived resources. Operationally, perceived stress can be measured by a person's self-report of adaptive capacity overload.

Self-Efficacy

When an individual perceives stress, based on pre-existing beliefs, values, and goals, he or she evaluates the threat (primary appraisal) and judges how much influence or effect they can have (secondary appraisal). Secondary appraisal involves the extent to which one believes they can influence or affect their own internal states, their behavior, the environment, and/or desired outcomes (perceived control) (Lazarus & Folkman, 1984; Wallston, Wallston, Smith & Dobbins, 1987). An essential component in judging opportunities for personal control in a stressful situation is the evaluation of one's self-efficacy with respect to implementing the required strategies (Bandura, 1997). Sometimes viewed as synonymous with behavioral control, as both refer to belief that one can perform a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

behavior, self-efficacy relates more to one's competence and to future behavior. Not a measure of the stress level itself, self-efficacy involves the belief that one is able to respond to demands of a stressful situation and it has both direct and indirect effects on health, including engaging in health-related behaviors (Schwarzer & Luszczynska, 2008).

Emotion regulation

Perceived stress and appraisal involve strong emotional reactions. Emotion regulation is a process by which a person can attempt to monitor and manage the experience and expression of their emotions and it can involve both conscious and unconscious processes (Gross & John, 2003). Difficulties in regulating emotions have been related to problems in psychological and physical functioning (Appleton, Buka, Loucks, Gilman, & Kubzansky, 2013; Gross & Munoz, 1995; DeSteno, Gross & Kubzansky, 2013). Eisenberg et al. (1997) used the term "self-regulation" to include both regulation of internal states and emotion-related behaviors. Coping can be considered under this classification when it involves efforts to decrease negative and increase positive emotions (and several coping measures include factors that involve emotion regulation), but it emphasizes primarily conscious processes as well as integration of cognitive and behavioral efforts (Folkman & Moskowitz, 2004). Despite the complexity involved in conceptualization and assessment, the study of emotion regulation can improve our understanding of how emotions can both help and hinder a person's attempts to master stressful situations (Cole, Martin & Dennis, 2004).

Coping

The most comprehensive and frequently-cited definition of coping was formulated by Lazarus and Folkman (1984): "cognitive and behavioral efforts to manage specific external or internal demands (and conflicts between them) that are appraised as taxing or exceeding the resources of the person." This definition emphasizes the importance of the context and the changing nature of coping in response to changes in the person-environment relationship. Operationally, coping can be measured as a person's self-report of the strategies he/she used to manage stressful situations. Among individuals with acute or chronic diseases, coping is a key factor that can determine how well an individual conceptualizes and behaviorally manages a health condition, and ways of coping with stress can influence whether individuals cope in healthy or health-destructive ways (Carver & Vargas, 2011; Folkman, 2010).

Based on the procedures described below, perceived stress, self-efficacy, emotion regulation and coping were initially considered to be the most important subdomains upon which to build a Toolbox measurement strategy.

The development of the overall NIH Toolbox project is described in Gershon et al., (2010; 2013) and the development of the Emotion domain is described in Salsman et al., 2013. The emotion domain was intended to include a wide spectrum of concepts related to emotion, including positive and negative emotion. Our process for identifying these subdomains was based on guidelines (Klem, Saghafi, Abromitis, Stover, Dew, & Pilkonis, 2009) and began with extensive literature reviews. An NIH Toolbox Request for Information (RFI) from experts in the field (Nowinski, et al., 2013) was followed by consultation with an

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

international group of content experts. These experts were identified through literature searches, the NIH CRISP database (now the NIH Research Portfolio Online Reporting Tools) or by nomination from a science office of one of the 12 institutes who were on the NIH Toolbox Project Team. Two hundred thirty-two experts were contacted, with a 64% response rate (N= 147). Of those, 61% whose response indicated expertise in emotional health received additional queries. Each respondent nominated four components of emotional health that would be most relevant for inclusion into the Toolbox. The criteria used for identifying the specific subdomains included relevance to health and disease, concurrence emerging from a series of interviews with content experts, and consultation with the emotional health domain team. Next, additional consultation with these experts was conducted to achieve consensus about essential subdomains. These considerations ultimately led to the development and refinement of conceptual definitions for the following emotional health sub-domains: negative affect, psychological well-being, stress & self-efficacy, and social relationships. Within each subdomain, variables considered to be essential were determined through additional consultation with experts, project staff and again, extensive literature reviews. In the Stress and Self-Efficacy domain, consensus was reached that the four relevant constructs identified were Perceived Stress, Self-Efficacy, Emotion Regulation, and Coping.

Method

Selection of Candidate Measures

Specialized teams were formed to focus on each of the sub-domains of emotional health and to develop criteria for evaluating existing measures that were potential candidates for inclusion in each subdomain. After achieving consensus on the inclusion of perceived stress, self-efficacy, emotion regulation, and coping, we sought recommendations for additional measures from experts and consultants and performed additional literature searches for measures to ensure content validity. A comprehensive literature search was conducted using the PubMed, PsycINFO, Buros Institute Test Reviews Online, Educational Testing Service, Patient-Reported Quality of Life Instrument Database (PROQOLID), Tests and Measures in the Social Sciences, and Health and Psychosocial Instruments (HAPI) databases to identify psychometrically-sound, free-for-use instruments. For the entire NIH Toolbox Emotional Health domain, 554 measures were identified, and of these, 127 were classified as potential measures of stress, self-efficacy, emotion regulation or coping. After review by project staff and emotion domain consultants, 44 measures were selected for further consideration based on criteria described earlier. For instruments with intellectual property restrictions that assessed the targeted areas, we contacted authors and/or publishers to request free-for-use status. We were granted permission to use 36 of those measures.

Criteria for selection in the initial phase included acceptable psychometric properties, including reliability, validity, availability of normative data, ability to use or be adapted across the lifespan, suitability for diverse populations, and versatility and portability in terms of types of studies (Gershon et al., 2013). In addition, because an objective of the NIH Toolbox was to characterize normative behavioral and neurological function, measures were excluded if they were developed and used primarily for identification or diagnosis of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

psychopathology. Another factor leading to exclusion of a measure was the existence of intellectual property issues that prevented free use in future research. A final criterion was that measures for each of the emotional health domain had to be brief enough so that respondents could complete the total Emotion Domain battery within 30 minutes (Gershon, et al., 2010; Salsman et al., 2013). The resulting short list included six measures that assessed stressful events or perceived stress, 3 that assessed self-efficacy, 4 that assessed emotion regulation and 31 that assessed coping.

The NIH Toolbox goal was to develop constructs that could be assessed throughout the lifespan. Recognizing the challenges in pediatric assessment the Pediatric consultant subgroup provided guidance in measurement selection and administration (Victorson, et al., 2013). We selected age-appropriate self-report items and measures for respondents ages 8 years and above and supplemented these self-report measures with proxy reports from parents or other caregivers for respondents ages 8–12 years. Parents and caregivers provided proxy data for coping and emotion regulation for children ages 3–7 years, and these data were collected and analyzed for initial calibration and validation. However, given the limited development of cognitive and emotional abilities for children ages 3–7 years and the inherent difficulty for parents or other caregivers to speculate on their children's perceptions of stress and self-efficacy, we did not assess these concepts for this age group. With further review of the short list, the expert panels, consultants and project staff selected the following measures to be included and assessed further for this subdomain.

Perceived stress—For the subdomain of perceived stress, The Perceived Stress Scale (PSS) (Cohen, Kessler & Gordon, 1995) was selected. The 14-item Likert format was administered. This scale is the most frequently used measure of a person's subjective experience of stress, including studies of health and quality of life, and has excellent reliability and validity (Cohen & Williamson, 1988, Monroe, 2008).

Self-Efficacy—The General Self-Efficacy Scale (GSES; Schwarzer & Jerusalem, 1995) scale was identified to measure general self-efficacy. It measures personal competence to deal effectively with a variety of stressful situations (Schwarzer & Luszczynska, 2008). It is a 10 item Likert scale and has excellent reliability, stability, and validity, and has been used in numerous studies related to health-related stress and behavior (Lusczynska, Gutierrez-Doña, & Schwarzer, 2005).

Emotion Regulation—Two scales of emotion regulation were also included in the initial calibration analyses. The Emotion Control Factor of the How I Feel measure (Walden, Harris, & Catron, 2003) is a 10-item scale that has established good reliability and validity. Three scales of the Children's Behavior Questionnaire (Low Intensity Pleasure, Inhibitory Control, and Soothability) were also included. The CBQ is a widely used measure of child development and behavior and has excellent psychometric properties ((Rothbart, Ahadi, Hershey, & Fisher, 2001).

Coping—Initial analyses were also conducted on several coping scales. The Ways of Coping Questionnaire (Folkman & Lazarus, 1988) is arguably the most widely used scale of coping strategies for adolescents and adults. It is a 50-item Likert scale with 8 empirically

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

derived factors. Another widely used scale of coping strategies is the Brief Cope (Carver, 1997), a 28 item scale with 14 2-item subscales with good psychometrics. For children and adolescents, two scales were initially considered: The Kidcope (Spirito, Stark & Williams, 1988), comprised of 10 items representing coping strategies and the Coping Scale for Children and Youth (CSCY) (Brodzinsky et al., 1992) comprised of 29 items with good psychometric properties.

As part of the NIH Toolbox project, all items were also reviewed by expert panels to identify and address potential issues related to assessment in special subgroups. Items were reviewed for relevance and use by older respondents (Geriatric Working Group), for cultural sensitivity and conceptual appropriateness across cultural groups (Cultural Working Group) and for Spanish translatability in the future (Spanish Language Working Group). These working groups reviewed all self-report items, identified potential administration difficulties, and, when necessary, offered alternative phrasing. For example, for adult measures, some items were revised so that the vocabulary was at or below a 6th grade reading level (for child self-reports, we used measures that were developed and validated specifically for this population). Victorson et al. (2013) provides further details of the working group review processes.

Sample/Participants

Initial testing of the identified measures was conducted by an internet survey company, Toluna (formerly Greenfield Online, http://www.greenfield.com). To recruit participants, Toluna sent email invitations to potential participants from their databases. Potential participants completed a rigorous screening process to determine their eligibility, which included validation of their internet protocol addresses to ensure that they were not fraudulently participating in surveys.. In agreeing to be participants in the Greenfield/Toluna online panels, participants had completed consent forms that applied to all future studies. Individual studies were explained to them in writing online. Consent was then assumed if they agreed to complete the measures. For the Toolbox studies all data were de-identified before being sent for instrument calibration and validation. In addition, all work in the study was approved by the Northwestern University Institutional Review Board. Following the initial screen, respondents completed a demographic survey, as well as the NIH Toolbox measures in the Emotion Domain. All participants who completed the survey were eligible for prize or incentive-based compensation through Toluna. Procedures for survey data quality control are described in detail at http://www.toluna-group.com/toluna-difference/data-quality/.

Five community-dwelling internet panel samples were recruited for the initial wave of testing (Total N=3,175). Subjects were drawn from the United States general population, and identified and recruited by Toluna,. These five samples were recruited based on age and consisted of adults aged 18 years and above (*N*= 1111), adolescents aged 13–17 years (*N*=512), and children aged 8–12 years (*N*= 513) for self-report forms; and children aged 8–12 years (*N*= 539) and children aged 3–7 years (*N*= 500) for parent/guardian proxy forms. Demographics for the five samples are provided in Table 1. This was a purposive sample designed with recruitment quotas for age, gender, and education. For calibration testing, an

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

internet panel was used to obtain a sufficient number of responses at each level, i.e. response option for all individual items.

Approach to data analyses – Initial Calibration Study

We collected self-report data in three different age groups, including 8–12, 13–17, and 18 years or older. In addition, we collected proxy-report data for 8–12 year-olds. Separate analyses were carried out for each of these groups. For sets of items that composed a single scale or subscale, we first examined their psychometric properties based on classical psychometric theory by examining Cronbach's alpha and corrected item-total correlations. We then used confirmatory factor analysis (CFA) to determine whether a one-factor model fit well, using cutoffs to guide to model interpretation. In cases for which we were unsure about dimensionality, we used CFA to verify known factor structures, and to examine the dimensionality of the various measures. In some cases where factor structures were unknown or not well-established, we relied on exploratory factor analyses.

For our CFAs, we assessed model fit using the comparative fit index (CFI) (Bentler, 1990; Hu & Bentler, 1999) and the root mean square error of approximation (RMSEA) (Browne, Cudeck, Bollen, & Long, 1993) to assess the fit of each model. We sought models with CFI >= .95 and RMSEA < .06, as suggested by Hu and Bentler (1999). We did not necessarily reject models outright if certain fit indices were slightly outside cutoffs (for a discussion of the use of cutoffs for fit indices, see Marsh, Hau, & Wen, 2004), and in some cases we accepted models for which RMSEA was below .10. We also examined differential item functioning (DIF), which determines whether an item has different psychometric characteristics in different groups. For each measure, we split the samples into two groups based on age (< 50 years-old vs. >= 50 years-old), gender, and education (one or more years of college vs. otherwise). For pediatric samples, we only compared participants on gender. We then examined each item across these groups using *lordif* – an algorithm developed by Choi, Gibbons, and Crane (2011) and implemented in the *R* statistical environment (see http://www.r-project.org/). Our DIF analyses flagged items using a criterion of pseudo-R^2 > .10 where a value of .20 corresponds to a small effect size.

Item response theory, which focuses on the relationship between performance on an item and a construct of interest, (Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991), was used to calibrate items on some of the measures in other emotional health subdomains of the NIH Toolbox. Because measures in this subdomain had sufficient psychometric properties, item response theory was not recommended for scoring because using the measures "as is" would allow for continuity with the extant literature. In the results that follow, information on the reliability and dimensionality of each measure is presented. We also present cases in which our analyses suggest concerns about, or modifications to, the measure.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Results

Perceived Stress

Across all age groups, internal consistency reliability was good for the Perceived Stress Scale (PSS) as indicated by Cronbach's alpha (see Table 2). One item (item #12) had low corrected item-total correlations across all groups: "In the last month, how often have you found yourself thinking about things that you have to accomplish?" In the adult sample, a one-factor model did not fit the PSS well, *CFI* = 0.88, *RMSEA* = .19. Because the PSS has some reversed-scored items, we suspected the presence of scoring factors, so we refit the model using a bi-factor approach (Holzinger & Swineford, 1937), in which one general factor is included along with one or more group factors to model local dependence. For the PSS, we included one group factor for all positively keyed items and another group factor for reverse-keyed items; this model achieved good fit for the PSS (see Table 2).

Analyses of children's data yielded similar results. Cronbach's alpha was .89 for 13–17 year-olds, .87 for 8–12 year-olds, and 0.87 for a proxy-report version of the PSS for 8–12 year-olds. Item #12 had a low item-total correlation in all three pediatric samples. Otherwise, the bi-factor model using scoring factors fit well in all three samples (see Table 3). The PSS was not administered to children below age 8 years. The scale authors (Cohen & Williamson, 1988) had also found a low factor loading for Item 12, as well as three other items and reduced the items to 10 (the PSS10). Their psychometric analyses found reliability and validity to be equivalent to the PSS14. After this analysis, we decided to recommend the PSS10 for the Toolbox.

Self-Efficacy

As shown in Table 3, the General Self Efficacy Scale (GSES) had high internal consistency for ages 8–12 years (for both self- and proxy-report), 13–17 years, and 18 years and older. One item had a low corrected item-total correlation (.16) in 8–12 year-olds: "If someone tries to keep me from getting what I want, I can find a way to get what I want." For the rest of the items, corrected item-total correlations were greater than .33 across all of the samples.

Our first attempt to fit a one-factor model was unsuccessful; *CFI* = .97 was acceptable, but *RMSEA* was >.10. A modification index indicated a residual dependency between two items with similar content (see Table 3), so the model was refit to account for it. This addition to the model resulted in a good-fitting one-factor model (see Table 4). This model also had good fit indices for adolescents, ages 8–12 years (self- and proxy report) and 13–17 years of age.

Emotion Regulation and Coping Scales—The Kidcope, CSCY, How I Feel-EC, Ways of Coping, Brief COPE, and CBQ had good reliability and validity according to classic psychometric analyses, but could not be fitted to a unidimensional CFA model. Given the criteria for NIH Toolbox selection, as well the results of these analyses, none of these scales were able to be included as Toolbox measures. One main reason was the lack of a single outcome score, as most were multidimensional measures and resulted in being too lengthy for inclusion. These measures have good psychometric properties, and, although

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

they were not included as Toolbox measures, they remain useful measures of the constructs they represent and were recommended as supplemental measures for research involving Toolbox measures.

Discussion

On the basis of a comprehensive literature review and extensive discussions, the NIH Toolbox emotional health team selected perceived stress, self-efficacy, emotion regulation and coping as important constructs to be evaluated for possible inclusion in the NIH Toolbox. To be selected for the NIH Toolbox, a measure had to have good psychometric properties, be comprised of a single factor, be free of intellectual property issues, and fit within a 30-minute test that also included measures of the other emotional subdomains of psychological well-being, negative affect, and social relationships. Thus, the measures that were judged most appropriate for NIH Toolbox purposes are widely used, psychometrically sound, and brief.

The preliminary analyses of the Perceived Stress Scale suggested that it measures a single dimension, although there was some variance accounted for by the direction of the items (positively- versus negatively-keyed). Based on our psychometric results, which were similar to those of the authors (Cohen & Williamson, 1988), we decided to recommend the 10-item version. Further testing of the PSS, including its association with other measures, will be completed as part of a national norming study. Because the model for the PSS included a general factor, it is possible that a computer-adaptive testing (CAT) protocol might be created to briefly and precisely measure perceived stress. Similarly, the one-factor model of the General Self-Efficacy Scale also had good psychometric properties and there is also the potential for CAT. Thus, these measures were deemed as appropriate for inclusion in this subdomain of the NIH Toolbox Emotional Health domain.

Model fit analyses of emotion regulation and general coping measures were less successful than those for stress and self-efficacy. Although disappointing because of the conceptual importance attributed to these measures in this subdomain, these findings are consistent with well-known complexities that characterize their assessment (Bridges, Denham & Ganiban, 2004; Folkman, & Moskowitz, 2004). For example, unlike other emotional health concepts like social support or life satisfaction, it is not conceptually feasible to derive a single summary score that adequately assesses the broad domain of coping. The key conceptual problem is that the evaluation of the adaptiveness/maladaptiveness of coping is determined in relation to a specific context. The key methodological problem is that coping is multi-dimensional with no agreement about the number of dimensions or a gold standard classification system (Skinner, Edge, Altman, & Sherwood, 2003). Because of these challenges and the constraints of the NIH Toolbox, these measures were judged unsuitable for this project, which may be a limitation of the NIH Toolbox.

In addition to these methodological issues, the study was limited by the participant selection process. First, by recruiting participants from an online research panel, we obtained samples that were predominantly White with more years of education relative to population-based samples. This may be evidence of a participation bias which is often unavoidable in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

volunteer research. Recent work, however, has shown that data collected from the Internet are comparable to data from probability-based general population samples (Liu et al., 2010). In addition, our scientific team has found that recruiting samples from Internet panels can be a cost-effective, efficient, and valid means of data collection as evidenced by our experience with two other large-scale NIH-funded efforts (Cella et al., 2010; Gershon, R. et al., 2012). Achieving representativeness was not a goal of this initial phase but will be addressed in the next phase with the norming sample.

This paper describes the first phase of instrument development of the Stress and Self-Efficacy sub-domain of the NIH Toolbox, an assessment battery for use in large-scale longitudinal epidemiologic and clinical trials studies. Consistent with past research, the Perceived Stress Scale and the General Self-Efficacy Scale had good psychometric properties in terms of dimensionality and the precision of scores. However, more information is needed on how these measures correlate with criteria of interest. Toward this end, a national norming study is planned that will provide further refinement of these measures, and the NIH Toolbox is slated to become a part of many future studies.

References

Antoni MH, Lutgendorf SK, Cole SW, Dhabhar FS, Sephton SE, McDonald PG, Sood AK. The influence of bio-behavioural factors on tumour biology: pathways and mechanisms. Nature Reviews. Cancer. 2006; 6(3):240–248. doi:10.1038/nrc1820. [PubMed: 16498446]

Appleton AA, Buka SL, Loucks Eb, Gilman SE, Kubzansky LD. Divergent associates of adaptive and maladaptive emotion regulation strategies with inflammation. Health Psychology. 2013; 32:748–756. doi:10.1037/a0030068. [PubMed: 23815767]

Bandura, A. Self-efficacy: the exercise of control. New York: W.H. Freeman; 1997.

Benight CC, Harper ML. Coping self-efficacy perceptions as a mediator between acute stress response and long-term distress following natural disasters. Journal of Traumatic Stress. 2002; 15(3):177–186. doi:10.1023/A:1015295025950. [PubMed: 12092909]

Bentler PM. Comparative fit indexes in structural models. Psychological Bulletin. 1990; 107(2):238–246. doi: 10.1037/0033-2909.107.2.238. [PubMed: 2320703]

Bridges LJ, Denham SA, Ganiban JM. Definitional issues in emotion regulation research. Child Development. 2004; 75:340–345. doi:10.1111/1467=8624:2004.00675.x. [PubMed: 15056188]

Brodzinsky DMJEM, Steiger C, Brodzinsky DM, Simon J, Gill M, Hitt JC. Coping scale for children and youth: Scale development and validation. Journal of Applied Developmental Psychology. 1992; 13(2):195–214. doi:10.1016/0193-3973(92)90029-H.

Browne, MW.; Cudeck, R.; Bollen, KA.; Long, KS. Testing Structural Equation Models. Newbury Park, CA: Sage Publications; 1993. Alternative ways of assessing model fit.

Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. International Journal of Behavioral Medicine. 1997; 4(1):92–100. http://link.springer.com/article/10.1207/s15327558ijbm0401_6. [PubMed: 16250744]

Carver, CS.; Vargas, S. Stress, coping, and health. In: Friedman, HS., editor. Oxford Handbook of Health Psychology. New York: Oxford University Press; 2011. p. 162-188.

Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S. PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. Journal of Clinical Epidemiology. 2010; 63(11):1179–1194. doi:10.1016/j.jclinepi.2010.04.011. [PubMed: 20685078]

Chesney MA, Neilands TB, Chambers DB, Taylor JM, Folkman S. A validity and reliability study of the coping self-efficacy scale. British Journal of Health Psychology. 2006; 11(Pt 3):421–437. doi: 10.1348/135910705X53155. [PubMed: 16870053]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Choi SW, Gibbons LE, Crane PK. 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. 2011; 39(8):1–30. doi: http://www.jstatsoft.org/v39/i08/. [PubMed: 21572908]

Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. JAMA. 2007; 298(14): 1685–1687. doi: 10.1001/jama298.14.1685. [PubMed: 17925521]

Cohen, S.; Kessler, RC.; Gordon, LU. Strategies for measuring stress in studies of psychiatric and physical disorder. In: Cohen, S.; Kessler, RC.; Gordon, LU., editors. Measuring stress : a guide for health and social scientists. New York: Oxford University Press; 1997. p. 3-26.

Cohen, S.; Williamson, GM. Perceived stress in a probability sample of the United States. In: Spacapan, S.; Oskamp, s, editors. The social psychology of health. Newbury Park, CA: Sage; 1988. p. 31-67.

Cole PM, Martin SE, Dennis TA. Emotion regulation as a scientific construct: methodological challenges and directions for child development research. Child Development. 2004; 75:317–333. doi:10.111/j.1467-8624.2004.00673.x. [PubMed: 15056186]

DeSteno D, Gross JJ, Kubzansky L. Affective states and health: the impact of emotion and emotion regulation. Health Psychology. 2013; 32:474–486. doi: 10.1037/a0030259. [PubMed: 23646831]

Eisenberg, N.; Fabes, RA.; Guthrie, IK. Coping with stress: the role of regulation and development. In: Wolchik, SA.; Sandler, IN., editors. Handbook of children's coping: linking theory and intervention. New York: Plenum; 1997. p. 41-70.

Embretson, SE.; Reise, SP. Item Response Theory for Psychologists. Mahwah, N.J.: Lawrence Erlbaum Associates; 2000.

Folkman S. Personal control and stress and coping processes: A theoretical analysis. Journal of Personality and Social Psychology. 1984; 46(4):839–852. http://psycnet.apa.org/journals/psp/46/4/839/. [PubMed: 6737195]

Folkman S. Stress, coping, and hope. Psycho-Oncology. 2010; 19(9):901–908. doi:10.1002/PON.1836. [PubMed: 20799373]

Folkman, S.; Lazarus, RS. Ways of coping questionnaire. Palo Alto, CA: Consulting Psychologists Press; 1988.

Folkman S, Moskowitz JT. Coping: pitfalls and promise. Annual Review of Psychology. 2004; 55:745–774. doi: 10.1146/annurev.psych.55.090902.141456.

Gershon R, Lai J, Bode R, Choi S, Moy C, Bleck T, Cella D. Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. Quality of Life Research. 2012; 21(3):475–486. doi: 10.1007/s11136-011-9958-8. [PubMed: 21874314]

Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, Wagster MV. Assessment of neurological and behavioural function: the NIH Toolbox. Lancet Neurology. 2010; 9(2):138–139. doi: S1474-4422(09)70335-7 [pii] 10.1016/S1474-4422(09)70335-7. [PubMed: 20129161]

Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH Toolbox for Assessment of Neurological and Behavioral Function. Neurology. 2013; 80(Suppl. 3):S2–S6. doi: 10.1212/WNL.o6013e31828772e5f. [PubMed: 23479538]

Gross JJ, John OP. Individual differences in emotion regulation processes: affective and social consequences. Journal of Personality and Social Psychology. 2003; 85:348–362. doi: 10.1037/0022-3514.85.2.346. [PubMed: 12916575]

Gross JJ, Munoz RF. Emotion regulation and mental health. Clinical Psychology Science and Practice. 1995; 2:151–164. doi: 10.1111/j.1468-2850.1995.tb00036.x.

Hambleton, RK.; Swaminathan, H.; Rogers, HJ. Fundamentals of Item Response Theory. Newbury Park, CA: SAGE Publications, Inc.; 1991.

Holzinger KJ, Swineford F. The bi-factor method. Psychometrika. 1937; 2:41–54.

Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6(1):1–55. doi: 10.1080/10705519909540118.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Klem M, Saghafi E, Abromitis R, Stover A, Dew M, Pilkonis P. Building PROMIS item banks: librarians as co-investigators. Quality of Life Research. 2009; 18(7):881–888. doi: 10.1007/s11136-009-9498-7. [PubMed: 19548118]

Lazarus, RS.; Folkman, S. Stress, appraisal, and coping. New York: Springer Publishing Co.; 1984.

Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, Hays RD. Representativeness of the PROMIS Internet Panel. Journal of Clinical Epidemiology. 2010; 63(11):1169–1178. doi:10.1016/j.jclinepi.2009.11.021. [PubMed: 20688473]

Luszczynska A, Gutierrez-Doña, Schwarzer R. General self-efficacy in various domains of human functioning: Evidence from five countries. International Journal of Psychology. 2005; 40:80–89. doi:10.1080/00207590444000041.

Marsh HW, Hau KT, Wen Z. In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Structural Equation Modeling. 2004; 11(3):320–341. doi:10.1207/s15328007sem1103_2.

Monroe SM. Modern approaches to conceptualizing and measuring human life stress. Annual Review of Clinical Psychology. 2008; 4:33–52. doi:10.1146/annurev.clinpsy.4.022007.141207.

Nowinski CJ, Victorson D, Debb SM, Gershon R. Input on NIH Toolbox Criteria: Surveying the End User Research Community. Neurology. 2013; 80 Supplement 3(11):S7–S12. doi:10.1212/WNL.0b013e3182872e4c. [PubMed: 23479548]

Raposa EB, Hammen CL, Brennan PA, O'Callaghan F, Najman JM. Early adversity and health outcomes in young adulthood: the role of ongoing stress. Health Psychology. 2014; 33:410–418. doi: 10.1037/a0032752. [PubMed: 23730716]

Rothbart MK, Ahadi SA, Hershey KL, Fisher P. Investigations of Temperament at Three to Seven Years: The Children's Behavior Questionnaire. Child Development. 2001; 72(5):1394–1408. doi: 10.1111/1467-8624.00355. [PubMed: 11699677]

Salsman JM, Butt Z, Pilkonis PA, Cyranowski JM, Zill N, Hendrie HC, Cella D. Emotion assessment using the NIH Toolbox. Neurology. 2013; 80 Supplement 3(11):S76–S86. doi: 10.1212/WNL.0b013e3182872e11. [PubMed: 23479549]

Schwarzer, R.; Jerusalem, M. Generalized Self-Efficacy Scale. In: Weinman, J.; Wright, S.; Johnston, M., editors. Measures in health psychology: A user's portfolio. Windsor, UK: NFER-Nelson; 1995. p. 35-37.

Schwarzer, R.; Luszczynska, A. Self-efficacy. Health Behavior Constructs: Theory, Measurement and Research. 2008. http://cancercontrol.cancer.gov/brp/constructs.

Segerstrom SC, Miller GE. Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. Psychological Bulletin. 2004; 130:601–630. doi: 10.1037/0033-2902.130.4.601. [PubMed: 15250815]

Skinner EA, Edge K, Altman J, Sherwood H. Searching for the structure of coping: a review and critique of category systems for classifying ways of coping. Psychological Bulletin. 2003; 129(2): 216–269. doi: 10.1037/0033-2909.129.2.216. [PubMed: 12696840]

Spirito A, Stark LJ, Williams C. Development of a brief coping checklist for use with pediatric populations. Journal of Pediatric Psychology. 1988; 13(4):555–574. doi:10.1093/jpepsy/13.4.555. [PubMed: 3216277]

Steptoe A, Kivimaki M. Stress and cardiovascular disease. Nature Reviews. Cardiology. 2012; 9(6): 360–370. doi: 10.1038/nrcardio.2012.45. [PubMed: 22473079]

Victorson D, Manly J, Wallner-Allen K, Fox N, Purnell C, Hendrie HC, Gershon RC. Using the NIH Toolbox in special populations: considerations for the assessment of pediatric, geriatric, culturally diverse, non-English speaking and disabled individuals. Neurology. 2013; 80 Supplement 3(11):S13–S19. doi:10.1212/WNL.0b013e3182872e26. [PubMed: 23479537]

Walden TA, Harris VS, Catron TF. How I Feel: A Self-Report Measure of Emotional Arousal and Regulation for Children. Psychological Assessment. 2003; 15:399–412. doi: 10.1037/1040-3590.5.3.399. [PubMed: 14593841]

Wallston KA, Wallston BS, Smith S, Dobbins CJ. Perceived control and health. Current psychological research and reviews. 1987; 6:5–25. http://link.springer.com/article/10.1007%2FBF02686633.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Demographic Characteristics for the NIH Toolbox Calibration and Validation Samples

	Pediatric Proxy 3–7 (*n* = 500)	Pediatric Proxy 8–12 (*n*= 539)	Pediatric 8–12 (*n*= 513)	Pediatric 13–17 (*n*=512)	Adults 18+ (*n*= 1,111)
Mean Age (SD)	5.0 (1.4)	10.1 (1.4)	10.1 (1.4)	15.1 (1.4)	45.4 (16.0)
	%	%	%	%	%
Sex					
Male	51	46	52	51	43
Ethnicity					
Not-Hispanic/Latino	88	89	89	94	93
Race					
American Indian/Alaska Native	1	1	<1	1	<1
Asian	2	2	2	1	5
Black/African-American	8	10	9	10	10
Native Hawaiian/Other Pacific Islander	1	<1	1	1	<1
White	77	79	81	83	80
Other/More than one	10	8	8	4	5
Proxy Relationship					
Mother/Female Guardian	76	71	N/A	N/A	N/A
Father/Male Guardian	17	23	N/A	N/A	N/A
Grandmother	7	6	N/A	N/A	N/A

Note: 61% of participants in the adult sample had some college education or more.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

The Perceived Stress Scale: Reliability and bifactor model fits.

Age Group	*N*	Cronbach's alpha	Description of model	CFI	RMSEA	Notes	DIF
18+	1111	.91	Bifactor model with scoring factors	.98	.09	Item #12 was included, but had a negative standardized factor loading (−.18) on the general factor.	None
13–17	512	.89	Bifactor model with scoring factors	.99	.06	Item #12 was included, but had a negative standardized factor loading (−.22) on the general factor.	None
8–12	513	.87	Bifactor model with scoring factors	.98	.07	Item #12 was included, but had a negative standardized factor loading (−.27) on the general factor.	None
8–12 Proxy-report	539	.87	Bifactor model with scoring factors	.99	.06	Item #12 was included, but had a negative standardized factor loading (−.25) on the general factor.	None

Note. DIF = Differential item functioning.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

General Self-Efficacy Scale: Reliability and model fits.

Age Group	N	Cronbach's alpha	Description of model	CFI	RMSEA	DIF
18+	1111	.93	One factor model with residual correlation between two items.*	.99	.07	None
13–17	512	.90	One factor model with residual correlation between two items.*	.99	.06	None for gender
8–12	513	.90	One factor model with residual correlation between two items.*	.99	.08	None for gender
8–12 proxy report	539	.92	One factor model with residual correlation between two items.*	.99	.07	One item was flagged for DIF across gender: "If your child is in trouble, he/she can usually think of a solution."

Note.

* The two items with a residual correlation were "I can always manage to solve difficult problems if I try hard enough." and "I can solve most problems if I invest the necessary effort."