Published in final edited form as:
J Clin Epidemiol. 2011 July ; 64(7): 794–804. doi:10.1016/j.jclinepi.2010.10.012.

Construction of the Eight Item PROMIS Pediatric Physical Function Scales: Built Using Item Response Theory

Esi Morgan DeWitt[1], **Brian D. Stucky**[2], **David Thissen**[2], **Debra E. Irwin**[3], **Michelle Langer**[4], **James W. Varni**[5], **Jin-Shei Lai**[6], **Karin B. Yeatts**[3], and **Darren A. DeWalt**[7]

[1] Department of Pediatrics, Duke University Medical Center, Durham, NC, USA

[2] Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[3] Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[4] National Board of Medical Examiners, Philadelphia, PA, USA

[5] Department of Pediatrics, College of Medicine, Department of Landscape Architecture and Urban Planning, College of Architecture, Texas A&M University, College Station, TX, USA

[6] Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

[7] Division of General Medicine and Clinical Epidemiology, Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Abstract

Objective—To create self-report physical function (PF) measures for children using modern psychometric methods for item analysis as part of Patient Reported Outcomes Measurement Information System (PROMIS).

Study Design and Setting—PROMIS qualitative methodology was applied to develop two PF item pools comprised of 32 mobility and 38 upper extremity items. Items were computer administered to subjects aged 8–17 years. Scale dimensionality and sources of local dependence (LD) were evaluated with factor analysis. Items were analyzed for differential item functioning (DIF) between genders. Items with LD, DIF, or low discrimination were considered for removal. Computerized adaptive testing performance was simulated, and short forms were constructed.

Results—3,048 children (51.8% female, 40% non-white, 22.7% chronically ill) participated. At least 754 respondents answered each item. Factor analytic results confirmed two dimensions of PF. Fifty-two of 70 items tested were retained. A 23 item mobility bank and a 29 item upper extremity bank resulted, and 8 item short forms were created. The item banks have high information from the population mean to 3 standard deviations below.

Conclusions—PROMIS pediatric PF item banks and 8-item short forms assess two dimensions, mobility and upper extremity function, and show good psychometric characteristics after large scale testing.

Address correspondence to: Esi Morgan DeWitt, MD MSCE, *now affiliated with* Division of Rheumatology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, MLC 4010; Cincinnati, OH, 45229. esi.morgan-dewitt@cchmc.org. Telephone: (513) 636-4676; Fax: (513) 636-4116.

Keywords

quality of life; outcome measure; disability; child; adolescent; psychometric methods

The Patient Reported Outcomes Measurement Information System (PROMIS) was created through a National Institutes of Health initiative to improve patient reported outcomes (PRO) assessment (1). PROMIS uses modern psychometric methods, including item response theory (IRT), to construct item banks from which static short forms or computerized adaptive tests (CAT) may be created to measure outcomes in a more efficient and precise manner than is possible using classical test theory(2). We describe the development of PROMIS physical function (PF) scales for pediatrics.

Item banks developed to satisfy the assumptions of IRT offer several advantages related to the measurement properties of IRT. Necessary conditions for item bank development are *unidimensionality*, that a scale measures a single underlying construct, lack of *local dependence* (LD), or that items share no covariance beyond that of the underlying construct, and lack of *differential item functioning* (DIF), meaning that people from different groups, (e.g., age, gender) who have a given level of an underlying trait, have the same probability of a given response. IRT based scales include the property of *interval level scaling* for better interpretation of change, calibration of items across a broad range of an underlying trait to overcome floor/ceiling effects, increased efficiency, and increased precision allowing more sensitivity to change(3). Furthermore, IRT based item banks support CAT, which employs an algorithm whereby only the most informative items targeting an individual's functioning levels are selected. CAT is in stark contrast to traditional, fixed- length questionnaires which, in order to capture a breadth of patient abilities, may result in patients answering items that are irrelevant to them and create high respondent burden.

There are examples of other disability scales developed using IRT, including the Activities Scale for Kids (ASK), for children with musculoskeletal disorders(4), and the Pediatric Evaluation of Disability Inventory (PEDI), for children with developmental disorders(5,6). The former includes domains of "personal care", "play", "locomotion", and others, while the latter divides PF into two dimensions, "mobility" and "self-care". Further, multi-dimensional CAT has been implemented in the PEDI(7). Yet, such measurement approaches have not been widely used outside of the disability community. The PROMIS scales aim to address the need for an IRT-based measurement system applicable across a range of health conditions, available for self- or proxy-administration, that is publicly available.

The PROMIS network aims to standardize PRO assessment across multiple chronic illness populations by creation of PRO item banks using a uniform methodology (8,9) to cover a range of domains of health-related quality of life (HRQOL). The framework for the health domains measured by PROMIS item banks is based on the WHO tripartite conceptualization of health (physical, social, and emotional)(1,10), with PF a central component of physical health. In addition to PF, PROMIS pediatric item banks were developed to measure pain, fatigue, anger, anxiety, depressive symptoms, peer relationships, and asthma symptoms by self-report in children ages 8–17 years old(11–14), with proxy-report versions in development for ages 5–17 years old. This report describes the construction and psychometric item analysis of the PROMIS pediatric physical function Mobility and Upper Extremity banks.

METHODS

The PROMIS pediatric PF domain was conceptualized as "one's ability to carry out various activities, ranging from self-care (activities of daily living) to more challenging and vigorous activities that require increasing degrees of mobility, strength, or endurance." We hypothesized that PF is multi-dimensional; in addition to the two dimensions considered in this project, other dimensions remain in need of measurement.

PROMIS methodology for initial item pool creation has been well described elsewhere(9). In brief, a multi-step process began with systematic identification and compilation of PRO items in existing scales. Items were then sorted by unidimensional aspects of a latent trait. New items were devised where there were apparent gaps in coverage across the continuum of a construct. After these processes, a pool of 177 candidate PF items was created. Redundant, vague, misclassified, confusing, or disease specific items were then set aside. Items selected for inclusion were re-written to conform to a standardized stem, recall period, and response options. Individual items were revised to reflect input from cognitive interviews(15). PF items have a standard 7-day recall period ("In the past 7 days"), are written in the past tense ("I could…"), and have a standard 5-point response option: "With no trouble, With a little trouble, With some trouble, With a lot of trouble, Not able to do." Items were written from the perspective of capability (16). Seventy PF items classified into two pools, Mobility (32) and Upper Extremity (38), remained for testing.

Sampling plan and recruitment

The items were divided across 4 different test forms, as listed in Tables 1a and 1b, and were computer administered to children aged 8–17 years. Each item was administered to at least 754 respondents. Due to a testing scheme that included several items from each of the item banks under development (e.g., PF, pain, fatigue), no one individual was administered the entire bank of PF items. Participants were recruited from medical clinics in North Carolina and Texas, and community schools in North Carolina. Parental informed consent and minor assent were obtained for all study participants. IRB approval was received from participating institutions.

The sample size requirements and testing scheme was designed to allow the following analytic plan: 1) assessing the domain factor structure, 2) tests for differential item function (DIF), 3) evaluation for local dependence, and 4) calibration of PF items using IRT methods.

Statistical and psychometric methods

The study population was characterized with descriptive statistics. We followed a standardized PROMIS framework for psychometric item analyses(8,12). Data quality was verified and analyses were conducted to ensure that IRT model assumptions were met. Item bank dimensionality was assessed with confirmatory factor analysis (CFA) of the inter-item polychoric correlation matrices using the DWLS algorithm in the computer program LISREL(17). Initial CFA models were fit with two correlated factors (Mobility and Upper Extremity).Investigations for LD included identifying significant error covariances between pairs or small clusters of items(18). If LD was identified only one of the items was selected from the subset to remain in the item bank and the others set aside.

Items in the banks which satisfied the unidimensionality criterion, as demonstrated by CFA, were subsequently calibrated using Samejima's Graded Response Model(19,20) using Multilog(21) software. For each item the GRM estimates a slope or discrimination parameter (a), which indicates the degree of association between the item responses and the underlying construct, in this case either Mobility or Upper Extremity, and four thresholds (b_k) (for five category items) that reflect the severity of physical functioning where the most

probable response occurs in a given category or higher. The fit of the IRT model was based on the *S-X* ^{2}statistic(22),(23), (24), in which a non-significant result is an indicator of adequate model fit.

DIF between males and females was evaluated using IRT-LR DIF as implemented in IRTLRDIF(25), (26)(the criterion of no DIF is a non-significant test-statistic). The Benjamini-Hockberg procedure was used to make inferential decisions in the context of the multiple comparisons(27),(28).

Short forms were created by selecting items from the calibrated item bank that were the most informative at one standard deviation below the mean (i.e., at T-score of 40). The appendix contains IRT scale scores computed for the summed scores of both short forms(29) as we expect the summed scores will be most useful for end-users.

RESULTS

The candidate pediatric PF items were administered to a racially diverse study cohort of 3,048 children. 22.7% had a chronic medical condition (Table 2).

Results from a two common factor model confirmed that there are two dimensions underlying the 70 item PF item pool, as proposed *a priori*, which we have labeled Mobility and Upper Extremity Function. However, the two dimensions proved to be highly correlated (from $r = 0.61$ to 0.93 across forms). Factor loadings and error covariances are displayed in Appendix Tables A1a through A1d. Using goodness of model fit indices as recommended by Reeve et al.(8), suggests adequate model fit. Form 1 (Table A1a), $\chi^2(72) = 64$, $p = 0.74$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00. Form 2 (Table A1b), χ^2 (100) = 118, p=0.10, CFI = 1.00, TLI = 1.00, RMSEA = 0.02. Form 3 (Table A1c), χ^2 (227) = 346, p = 0.00, CFI = 1.00, TLI = 1.00, RMSEA = 0.03. Form 4 (Table A1d), χ^2 (204) = 288, p=0.00 CFI = 1.00, TLI = 0.99, RMSEA = 0.03.

On each of the four forms there were doublets or triplets of items which exhibited LD. This can happen when subsets of items share content or wording that is similar, yet different from the scale's other items. For example, Form 4 has three locally dependent mobility items, "I could walk a mile", "I could walk more than one block" and "I could keep up when I played with other kids" (Appendix Table A1d). These all convey the shared idea of endurance, quite distinct from other items on the form. In order to ensure unidimensionality of the scales, in general only one item from each doublet or triplet was preserved in the final item bank (Appendix Tables A1a-d; A2a, b). Two exceptions were in Form 1, of two Upper Extremity items with LD, both were excluded due to poor psychometric performance. Two Mobility items that exhibited statistical LD to a small (albeit significant) extent were retained in the final item pool. Inclusion/exclusion decision was made via team discussion with an attempt to take both clinical and psychometric perspectives into account.

Eight items (6 Mobility and 2 Upper Extremity) were set aside after the factor analyses, due to one or more of the following reasons: loading on both factors, LD or low relationship with the construct or too little variation. These eight items are marked with the notation "(d)" in Tables A1a-d. In Form 3, one of the Mobility items, "I needed help with a bath" loaded onto the Upper Extremity Factor, and was thus moved to the Upper Extremity item pool.

After the factor analyses were complete, locally independent subsets of items were calibrated using the GRM. To avoid calibrating LD items, each calibration included only a single item from each LD doublet or triplet. Appendix tables A2a and A2b show the item parameter estimates sorted based on the magnitude of information each item provides at one

standard deviation below the mean. The tables also provide item fit ($S\text{-}X^2$) and DIF statistics ($LR\ X^2$) for the final item banks. Listed at the bottom of the tables are the 10 additional items that were set aside during the final assembly of the pool. Three items set aside for LD or poor model fit, five items for DIF, and two items for their lack of discrimination. The final Mobility item bank contained 23 items and the final Upper Extremity Bank contained 29 items.

Information functions of the full item bank and short form for Mobility and Upper Extremity are shown in Figure 1 and Figure 2, respectively. These are presented on a *T*-score scale in which the mean is set at 50 with a standard deviation of 10. Test information is the expected value of the inverse of the squared standard error of measurement, such that a standard error of measurement of approximately .32 (or 3.2 on a *T*-score metric) is associated with a test information value of 10, which corresponds to a reliability coefficient of 0.90. The short-forms for both Mobility and Upper Extremity (selected as the 8 most informative items at one standard deviation below population mean (i.e., at 40 on a *T*-score metric)) have information values greater than 10 between scores of about 20 to 45 for Mobility and 20 to 40 for Upper Extremity.

Figures 1 and 2 also function as simulated CATs. Both figures contain five information functions which represent the collective information of 8 (potentially different) items that individuals at five different score locations (30 through 70, on a *T*-score scale) would receive given a perfectly operating CAT. In this case, because the items are informative in generally the same locations, a CAT may not dramatically improve the efficiency of the questionnaire. Nonetheless, PROMIS Assessment Center (http://www.assessmentcenter.net/) contains the calibrated item bank and allows the user to select and administer items as a CAT.

DISCUSSION

PROMIS pediatric PF item banks and short forms assessing two dimensions of PF, Mobility and Upper Extremity Function, show strong psychometric characteristics after initial large scale testing. Factor analysis supported the creation of separate PF item banks. While the complete set of items in either item bank was not tested by individuals due to concerns over respondent burden (a potential limitation), an advantage of this approach is that we had replication of the findings across four forms, providing helpful evidence that persuaded us to keep the banks separate. Although the item banks represent different dimensions of PF they are not-surprisingly, highly correlated. The approach to measurement of PF in the PROMIS pediatric item bank we present here diverges from that of the PROMIS PF bank for adult patients, in which PF is treated as a unidimensional construct, though not without debate(30).

With few exceptions(31),(32), traditional PF scales do not allow disaggregation of various aspects of PF, such as lower or upper body function. Aggregation of multiple aspects of PF into a single summary score may blunt the instruments, i.e., reduce precision, mute responsiveness to change, and on a more basic level reduce information and interpretability. The realm of PF assessment and how to define and handle its apparent subdomains is an underdeveloped area, resulting in divergent approaches to its measurement. Although multidimensional CAT is attractive in the potential for increased efficiency of measurement, this method requires added complexity in scoring(33).

One of the primary advantages of IRT derived measurement scales is the potential to overcome floor/ceiling effects by broadening the range of measurement. Despite following PROMIS procedures for item development and testing, the PROMIS Pediatric PF scales

show a ceiling effect. Indeed the scaled scores for Mobility and Upper Extremity Function short forms reach only 59 and 57, respectively, less than one standard deviation above the population mean, with a broader range at lower levels of function, down to 14 and 10, respectively.

Large scale testing is currently underway in children with chronic illnesses, including cancer, chronic kidney disease, and rheumatic diseases. Future work is needed for translation and cross cultural validation. Additional item development is needed to enhance measurement of individuals with levels of function above the population mean.

Summary

We described the item development, item analyses, and construction of the first version of the PROMIS Pediatric Mobility and Upper Extremity Function item banks. These instruments were tested in a large, diverse population of children ages 8–17 years and show excellent test properties in preliminary testing. Additional work is underway to further validate and calibrate the instruments in a variety of chronic illness populations.

Acknowledgments

We are grateful to Harry A. Guess, MD, PhD, under whose vision and leadership this PROMIS project to develop item banks for pediatrics took shape.

Funding for this research was provided to participating institutions by the National Institutes of Health through the NIH Roadmap for Medical Research, Cooperative Agreements 1U01AR052181-01 to University of North Carolina, PI: Darren DeWalt, MD, MPH; and U01AR52186 to Duke University, PI: Kevin Weinfurt, PhD. Additional information on the Patient Reported Outcomes Measurement Information System is available at http://nihroadmap.nih.gov and http://www.nihpromis.org

References

1. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care. 2007 May; 45(5 Suppl 1):S3–S11. [PubMed: 17443116]
2. Flynn KE, Dombeck CB, DeWitt EM, Schulman KA, Weinfurt KP. Using item banks to construct measures of patient reported outcomes in clinical trials: investigator perceptions. Clin Trials. 2008; 5(6):575–86. [PubMed: 19029206]
3. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual Life Res. 2007; 16(Suppl 1):5–18. [PubMed: 17375372]
4. Young NL, Williams JI, Yoshida KK, Wright JG. Measurement properties of the activities scale for kids. J Clin Epidemiol. 2000 Feb; 53(2):125–37. [PubMed: 10729684]
5. Haley SM, Raczek AE, Coster WJ, Dumas HM, Fragala-Pinkham MA. Assessing mobility in children using a computer adaptive testing version of the pediatric evaluation of disability inventory. Arch Phys Med Rehabil. 2005 May; 86(5):932–9. [PubMed: 15895339]
6. Mulcahey MJ, Haley SM, Duffy T, Pengsheng N, Betz RR. Measuring physical functioning in children with spinal impairments with computerized adaptive testing. J Pediatr Orthop. 2008 Apr–May; 28(3):330–5. [PubMed: 18362799]
7. Haley SM, Ni P, Ludlow LH, Fragala-Pinkham MA. Measurement precision and efficiency of multidimensional computer adaptive testing of physical functioning using the pediatric evaluation of disability inventory. Arch Phys Med Rehabil. 2006 Sep; 87(9):1223–9. [PubMed: 16935059]
8. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care. 2007 May; 45(5 Suppl 1):S22–31. [PubMed: 17443115]

9. DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: the PROMIS qualitative item review. Med Care. 2007 May; 45(5 Suppl 1):S12–21. [PubMed: 17443114]
10. WHO. World Health Organization Technical Report Series No. 137. Geneva, Switzerland: World Health Organization; 1957. Measurement of Levels of Health: Report of a Study Group.
11. Irwin DE, Stucky B, Langer MM, Thissen D, Dewitt EM, Lai JS, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. Qual Life Res. May; 19(4): 595–607. [PubMed: 20213516]
12. Yeatts K, Stucky B, Thissen D, Irwin DE, Varni JW, DeWitt EM, et al. Construction of the Pediatric Asthma Impact Scale (PAIS) for the Patient Reported Outcomes Measurement Information System (PROMIS). Journal of Asthma. in press.
13. Varni JW, Stucky BD, Thissen D, Dewitt EM, Irwin DE, Lai JS, et al. PROMIS Pediatric Pain Interference Scale: An Item Response Theory Analysis of the Pediatric Pain Item Bank. J Pain. Jun 1.
14. Irwin DE, Stucky BD, Thissen D, Dewitt EM, Lai JS, Yeatts K, et al. Sampling plan and patient characteristics of the PROMIS pediatrics large-scale survey. Qual Life Res. May; 19(4):585–94. [PubMed: 20204706]
15. Irwin DE, Varni JW, Yeatts K, DeWalt DA. Cognitive interviewing methodology in the development of a pediatric item bank: a patient reported outcomes measurement information system (PROMIS) study. Health Qual Life Outcomes. 2009; 7:3. [PubMed: 19166601]
16. Chakravarty EF, Bjorner JB, Fries JF. Improving Patient Reported Outcomes Using Item Response Theory and Computerized Adaptive Testing. J Rheumatol. 2007; 34:1426–31. [PubMed: 17552069]
17. Joreskog, KG.; Sorbom, D. LISREL 8.5. Lincolnwood, IL: Scientific Software International Inc; 2003.
18. Hill CD, Edwards MC, Thissen D, Langer MM, Wirth RJ, Burwinkle TM, et al. Practical issues in the application of item response theory: a demonstration using items from the pediatric quality of life inventory (PedsQL) 4.0 generic core scales. Med Care. 2007 May; 45(5 Suppl 1):S39–47. [PubMed: 17443118]
19. Samejima F. Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph. 1969; (17)
20. Samejima, F. Graded Response Model. In: van der Linden, WJ.; Hambleton, RK., editors. Handbook of Modern Item Response Theory. New York: Spinger; 1997. p. 85-100.
21. du Toit, M., editor. IRT from SSI: Bilog-MG, Multilog, Parscale, Testfact. Lincolnwood, IL: Scientific Software International; 2003.
22. Orlando M, Thissen D. LIkelihood-based item-fit indices for dichotomous item response theory models. Applied Psychological Measurement. 2000; 24:50–64.
23. Orlando M, Thissen D. Further Examination of the Performance of S-X2, an item fit index for dichotomous item response theory models. Applied Psychological Measurement. 2003; 27:289–98.
24. Bjorner, JB.; Smith, KJ.; Edelen, MO.; Stone, C.; Thissen, D.; Sun, X. IRTFIT: A Macro for Item Fit and Local Dependence Tests under IRT Models. Lincoln, RI: QualityMetric Incorporated; 2007.
25. Thissen, D.; Steinberg, L.; Wainer, H. Detection of Differential Item Functioning Using the Parameters of Item Response Models. In: Holland, PW.; Wainer, H., editors. Differential Item Functioning. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993. p. 67-113.
26. Thissen, D. IRTLRDIF: Software for the Computation of the Statistics Involved in Item Response Theory Likelihood-Ratio Test for Differential Item Functioning. Chapel Hill, NC: LL Thurstone Psychometric Laboratory, The University of North Carolina at Chapel Hill; 2001.
27. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B. 1995; 57:289–300.
28. Williams VSL, Jones LV, Tukey JW. Controlling Error in Multiple Comparisons, with Examples from State-to-State Differences in Educational Achievement. J Educ Behav Stat. 1999; 24:42–69.

29. Thissen, D.; Nelson, L.; Rosa, K.; McLeod, LD. Item Response Theory for Items Scored in More than Two Categories. In: Thissen, D.; Wainer, H., editors. Test Scoring. Mahwah, NJ: Lawrence Erlbaum Associates; 2001.

30. Rose M. J Clin Epidemiol. 2008; 61:17–33. [PubMed: 18083459]

31. Filocamo G, Sztajnbok F, Cespedes-Cruz A, Magni-Manzoni S, Pistorio A, Viola S, et al. Development and validation of a new short and simple measure of physical function for juvenile idiopathic arthritis. Arthritis Rheum. 2007 Aug 15; 57(6):913–20. [PubMed: 17665481]

32. Pruitt SD, Seid M, Varni JW, Setoguchi Y. Toddlers with limb deficiency: conceptual basis and initial application of a functional status outcome measure. Arch Phys Med Rehabil. 1999 Jul; 80(7):819–24. [PubMed: 10414768]

33. Reckase, MD. Multidimensional Item Response Theory. New York, NY: Springer; 2009.

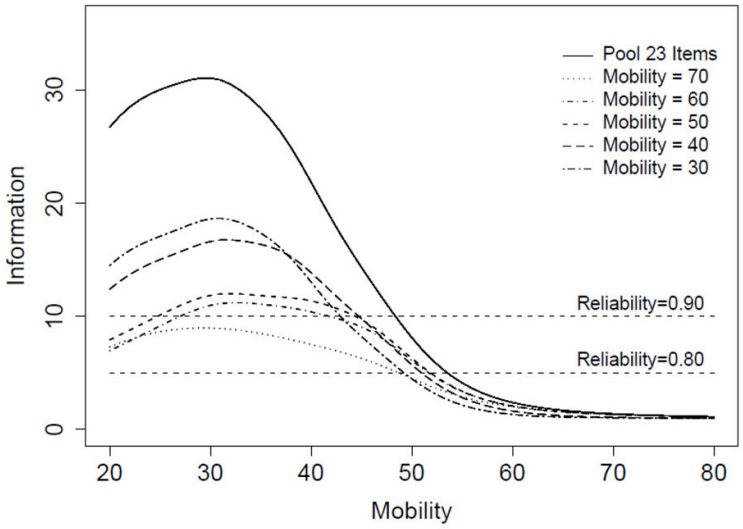

Figure 1.
Mobility Test Information Curves. Test information curves are displayed for the 23 item mobility pool, and for the subsets of eight items from the pool that are most informative at *T*-scores of 30, 40, 50, 60, and 70.

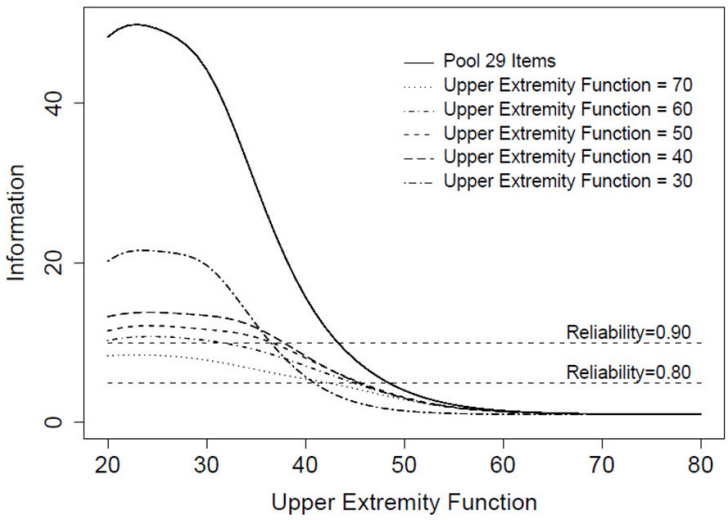

Figure 2.
Upper Extremity Test Information Curves. Test information curves are displayed for the 29 item upper extremity pool, and for the subsets of eight items from the pool that are most informative at *T*-scores of 30, 40, 50, 60, and 70.

Table 1a

PROMIS Pediatric Physical Function Mobility item stems administered on 4 Forms for item analysis and calibration

Form 1	Form 2	Form 3	Form 4
I have been physically able to do the activities I enjoy most.	I could run a mile.	I could walk.	I could walk a mile.
I could ride a bike.	I could run.	I could get into bed by myself.	I could get in and out of a car.
I could do sports and exercise that other kids my age could do.	I could do sports activity or exercise.	I could stand up by myself.	I could walk more than one block.
I could take a bath by myself.	I could walk up stairs without holding on to anything.	I needed help with a bath.	I could keep up when I played with other kids.
I could get down on my knees without holding on to something.	I could keep my balance.	I used a walker, cane or crutches to get around.	I could take a shower by myself.
I could go up one step.	I could get up from a regular toilet.	I could move my legs.	I could get out of bed by myself.
	I could stand up on my tiptoes.	I could get up from the floor.	I used a wheelchair to get around.
		I could turn my head all the way to the side.	I could carry my books in my backpack.
		I could run fast.	I could bend over to pick something up.
		I could walk across the room.	

Table 1b

PROMIS Pediatric Physical Function Upper Extremity item stems administered on 4 Forms for item analysis and calibration

Form 1	Form 2	Form 3	Form 4
I could tie shoelaces by myself.	I could move my hands or fingers.	I could lift something heavy.	I could write with a pen or pencil.
I could put on my clothes by myself.	I could put on my shoes by myself.	I could put on socks by myself.	I could brush my teeth by myself.
I could hold an empty cup.	I could button my shirt or pants.	I could put toothpaste on my toothbrush by myself.	I could turn door handles by myself.
I could pull on and fasten my seatbelt.	I could use a mouse or touch pad for the computer.	I could pull a shirt on over my head by myself.	I could put on my pants by myself.
I could open a bag of chips.	I could lift a cup to drink.	I could hold a full cup.	I could use a keyboard on the computer.
I could open plastic food containers.	I could undo Velcro.	I could zip up my clothes.	I could cut my food.
	I could cut paper with scissors.	I could use a key to unlock a door.	I could wipe myself after using the toilet.
	I could wash my face with a cloth.	I could dial a phone.	I could open my clothing drawers.
		I could pull open heavy doors.	I could pour a drink from a full pitcher.
		I could turn pages in a book.	I could carry a tray with food on it.
		I could open the rings in school binders.	I could dry my back with a towel.
		I used a pencil with a special grip to write.	

Table 2

Study subject characteristics

	Form 1 n=759(%)	Form 2 n=770 (%)	Form 3 n=754 (%)	Form 4 n=765 (%)	Total Forms 1 to 4 n= 3,048 (%)
Child's Gender					
Male	382 (50.3)	351 (45.6)	355 (47.1)	382 (49.9)	1,470 (48.2)
Female	377 (49.7)	419 (54.4)	399 (52.9)	383 (50.1)	1,578 (51.8)
Missing	0	0	0	0	0
Child's Age (yrs)					
8–12	446 (58.8)	441 (56.4)	303 (40.2)	426 (55.7)	1,616 (53.0)
13–17	312 (41.1)	326 (42.3)	451 (59.8)	337 (44.0)	1,426 (46.8)
Missing	1 (0.1)	3 (0.3)	0	2 (0.3)	6 (0.2)
Child's Race					
White	457 (60.2)	452 (58.7)	457 (60.6)	462 (60.4)	1,828 (60.0)
Black or African-American	154 (20.2)	168 (21.8)	172 (22.8)	150 (19.6)	644 (21.1)
American Indian/Alaska Native	5 (0.6)	10 (1.3)	7 (0.9)	10 (1.3)	32 (1.0)
Asian	12 (1.6)	13 (1.7)	6 (0.8)	10 (1.3)	41 (1.3)
Native Hawaiian Other Pacific Is	0	1 (0.1)	2 (0.3)	2 (0.3)	5 (0.2)
Other	58 (7.6)	50 (6.5)	58 (7.7)	64 (8.4)	230 (7.5)
Multiple Races	47 (6.2)	54 (7.0)	27 (3.6)	43 (5.6)	171 (5.6)
Missing	26 (3.4)	22 (2.9)	25 (3.3)	24 (3.1)	97 (3.2)
Child's Ethnicity					
Non Hispanic	614 (80.9)	641 (83.2)	617 (81.8)	619 (80.9)	2,491 (81.7)
Hispanic	141 (18.6)	121 (15.7)	131 (17.4)	141 (18.4)	534 (17.5)
Missing	4 (0.5)	8 (1.1)	6 (0.8)	5 (0.7)	23 (0.8)
Child's Chronic Conditions Past 6 mo					
No	600 (79.0)	580 (75.3)	569 (75.5)	592 (77.4)	2,341 (76.8)
Yes	157 (20.7)	187 (24.3)	180 (23.9)	169 (22.1)	693 (22.7)
Missing	2 (0.3)	3 (0.4)	5 (0.6)	4 (0.5)	14 (0.5)
Guardian's* Relationship to Child					
Parent	696 (91.7)	717 (93.1)	695 (92.2)	708 (92.6)	2,816 (92.4)
Grandparent	32 (4.2)	30 (3.9)	32 (4.2)	43 (5.6)	137 (4.5)

	Form 1 n=759(%)	Form 2 n=770 (%)	Form 3 n=754 (%)	Form 4 n=765 (%)	Total Forms 1 to 4 n= 3,048 (%)
Guardian or Other	31 (4.1)	21 (2.7)	26 (3.5)	13 (1.7)	91 (3.0)
Missing	0	2 (0.3)	1 (0.1)	1 (0.1)	4 (0.1)
Guardian's* Education Level					
<= 8th grade	12 (1.6)	16 (2.3)	13 (1.8)	16 (2.1)	57 (1.9)
Some high school	39 (5.1)	34 (4.4)	54 (7.2)	55 (7.2)	182 (6.0)
High school degree/GED	151 (19.9)	153 (19.7)	163 (21.6)	159 (20.8)	626 (20.5)
Some college/technical degree	255 (33.6)	245 (31.8)	251 (33.2)	260 (34.0)	1011 (33.2)
College degree	179 (23.6)	214 (27.8)	183 (24.3)	180 (23.5)	756 (24.8)
Advanced degree	121 (15.9)	105 (13.6)	86 (11.4)	95 (12.4)	407 (13.4)
Missing	2 (0.3)	3 (0.4)	4 (0.5)	0	9 (0.3)
Data Collection Site					
Schools – NC	57 (7.5)	57 (7.4)	49 (6.5)	51 (6.7)	214 (7.0)
Clinics - NC	349 (46.0)	350 (45.5)	343 (45.5)	351 (45.9)	1,393 (45.7)
Clinics – TX	353 (46.5)	363 (47.1)	362 (48.0)	363 (47.4)	1,441 (47.3)

*guardian, parent or caregiver completing sociodemographic form and signing consent documents