PROSETTA STONE® ANALYSIS REPORT

A ROSETTA STONE FOR PATIENT REPORTED OUTCOMES

NEURO-QOL PEDIATRIC COGNITIVE FUNCTION AND PEDIATRIC PERCEIVED COGNITIVE FUNCTION

DAVID CELLA , BENJAMIN D. SCHALET, MICHAEL A. KALLEN, JIN-SHEI LAI, KARON F. COOK, JOSHUA RUTSOHN & SEUNG W. CHOI

DEPARTMENT OF MEDICAL SOCIAL SCIENCES
FEINBERG SCHOOL OF MEDICINE
NORTHWESTERN UNIVERSITY

This research was supported by an NIH/National Cancer Institute grant PROSETTA STONE (1RC4CA157236-01, PI: David Cella). Authors acknowledge careful reviews, comments, and suggestions from Drs. Robert Brennan, Lawrence Hedges, Won-Chan Lee, and Nan Rothrock.

Table of Contents

1. Introduction 3
2. The PRO Rosetta Stone Project 3
2.1. Patient-Reported Outcomes Measurement Information System (PROMIS) 4
2.2. The NIH Toolbox for Assessment of Neurological and Behavioral Function (NIH Toolbox) 5
2.3. Quality of Life Outcomes in Neurological Disorders (Neuro-QoL) 5
3. Legacy Instruments 6
3.1. Pediatric Perceived Cognitive Function Item Bank (Peds PCF) 6
4. Linking Methods 6
4.1. IRT Linking 7
4.2. Equipercentile Linking 8
4.3. Linking Assumptions 9
5. Linking Results 10
5.8 Neuro-QoL Pediatric Cognitive Function and Pediatric Perceived Cognitive Function 11
5.8.1 Raw Summed Score Distribution 11
5.8.2 Classical Item Analysis 12
5.8.3 Confirmatory Factor Analysis (CFA) 12
5.8.4 Item Response Theory (IRT) Linking 13
5.8.5 Raw Score to T-Score Conversion using Linked IRT Parameters 16
5.8.6 Equipercentile Linking 16
5.8.7 Summary and Discussion 17
6. Appendix Table 21: Raw Score to T-Score Conversion Table (IRT Fixed Parameter Calibration Linking) for Pediatric PCF 30-Item Short Form and Neuro-QoL Pediatric Cognitive Function (PROsetta Stone Wave 3 Study) - RECOMMENDED 20
7. Appendix Table 22: Direct (Raw to Scale) Equipercentile Crosswalk Table – From Pediatric PCF 30-Item Short Form to Neuro-QoL Pediatric Cognitive Function – Table 21 is recommended 22
8. Appendix Table 23: Indirect (Raw to Raw to Scale) Equipercentile Crosswalk Table – From Pediatric PCF 30-Item Short Form to Neuro-QoL Pediatric Cognitive Function – Table 21 is recommended 25
9. Pediatric Perceived Cognitive Function 30-Item Short Form 28

PRO Rosetta Stone (*PROsetta Stone®*) Analysis

1. Introduction

A common problem when using a variety of patient-reported outcome measures (PROs) for diverse populations and subgroups is establishing the comparability of scales or units on which the outcomes are reported. The lack of comparability in metrics (e.g., raw summed scores vs. scaled scores) among different PROs poses practical challenges in measuring and comparing effects across different studies. Linking refers to establishing a relationship between scores on two different measures that are not necessarily designed to have the same content or target population. When tests are built in such a way that they differ in content or difficulty, linking must be conducted in order to establish a relationship between the test scores. One technique, commonly referred to as equating, involves the process of converting the system of units of one measure to that of another. This process of deriving equivalent scores has been used successfully in educational assessment to compare test scores obtained from parallel or alternate forms that measure the same characteristic with equal precision. Extending the technique further, comparable scores are sometimes derived for measures of different but related characteristics. The process of establishing comparable scores generally has little effect on the magnitude of association between the measures. Comparability may not signify interchangeability unless the association between the measures approaches unit reliability. Equating, the strongest form of linking, can be established only when two tests 1) measure the same content/construct, 2) target very similar populations, 3) are administered under similar conditions such that the constructs measured are not differentially affected, 4) share common measurement goals and 5) are equally reliable. When test forms are created to be similar in content and difficulty, equating adjusts for differences in difficulty. Test forms are then considered to be essentially the same, so scores on the two forms can be used interchangeably after equating has adjusted for differences in difficulty. For tests with lesser degrees of similarity, only weaker forms of linking are meaningful, such as calibration, concordance, projection, or moderation.

2. The PRO Rosetta Stone Project

The primary aim of the PRO Rosetta Stone (PROsetta Stone®) project (1RC4CA157236-01, PI: David Cella) is to develop and apply methods to link the Patient-Reported Outcomes Measurement Information System (PROMIS) measures with other related “legacy” instruments in order to expand the range of PRO assessment options within a common, standardized metric. The project identifies and applies appropriate linking methods that allow scores on a range of legacy PRO instruments to be expressed as standardized T-score metrics linked to the PROMIS metric. This report (Volume 3) encompasses 8 linking studies based on available pediatric PRO data from NIH Toolbox, Neuro-QoL, and PROsetta Stone Wave 3. The PROsetta Stone Report Volume 1 included linking results primarily from PROMIS Wave 1, as well as links based on NIH Toolbox and Neuro-QoL data. Volume 2 included linking studies based on data that were primarily from PROsetta Stone Waves 1 and 2.

2.1. Patient-Reported Outcomes Measurement Information System (PROMIS)

In 2004, the NIH initiated the PROMIS[1] cooperative group under the NIH Roadmap[2] effort to re-engineer the clinical research enterprise. The aim of PROMIS is to revolutionize and standardize how PRO tools are selected and employed in clinical research. To accomplish this, a publicly-available system was developed to allow clinical researchers access to a common repository of items and state-of-the-science computer-based methods for administering the PROMIS measures. The PROMIS measures include item banks across a wide range of domains that comprise physical, mental, and social health for adults and children, with 12-124 items per bank. Initial concepts measured include emotional distress (anger, anxiety, and depression), physical function, fatigue, pain (quality, behavior, and interference), social function, sleep disturbance, and sleep-related impairment. The banks can be used to administer computerized adaptive tests (CAT) or fixed-length forms in these domains. We have also developed 4-item to 20-item short forms for each domain, and a 10-item Global Health Scale that includes global ratings of five broad PROMIS domains and general health perceptions. As described in a full issue of *Medical Care* (Cella et al., 2007), the PROMIS items, banks, and short forms were developed using a standardized, rigorous methodology that began with constructing a consensus-based PROMIS domain framework.

All PROMIS banks have been calibrated according to Samejima's (1969) graded response model and are based on large data collections including both general and clinical samples. All PROMIS banks are re-scaled (mean=50 and SD=10) using scale-setting subsamples matching the marginal distributions of gender, age, race, and education in the 2000 US census. The PROMIS Wave I calibration data included (a) a small number of full-bank testing cases (approximately 1,000 per bank) from a general population taking one full bank and (b) a larger number of block-administration cases (n= ~14,000) from both general and clinical populations taking a collection of blocks representing all banks, with seven items administered from each bank. The full-bank testing samples were randomly assigned to one of seven different forms. Each form was composed of one or more PROMIS domains (with an exception of Physical Function, where the bank was split over two forms) and one or more legacy measures of the same or related domains.

The PROMIS Wave I data collection design included a number of widely accepted "legacy" measures. The legacy measures used for validation evidence included Buss-Perry Aggression Questionnaire (BPAQ), Center for Epidemiological Studies Depression Scale (CES-D), Mood and Anxiety Symptom Questionnaire (MASQ), Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F), Brief Pain Inventory (BPI), and SF-36. In addition to PROMIS-legacy measure pairings for validity assessment (e.g., PROMIS Depression and CES-D), the PROMIS Wave I data allowed for the potential to link over a dozen pairs of measures/subscales. Furthermore, included within each of the PROMIS banks were items from many other existing measures. Depending on the nature and strength of relationship between the measures, various linking procedures can be used to allow for cross-walking of scores.

[1] www.nihpromis.org
[2] www.nihroadmap.nih.gov

(Note that most of the linking reports based on the PROMIS Wave 1 dataset are included in Volume 1.)

2.2. The NIH Toolbox for Assessment of Neurological and Behavioral Function (NIH Toolbox)

Developed in 2006 with the NIH Blueprint funding for Neuroscience Research, four domains of assessment central to neurological and behavioral function were created to measure cognition, sensation, motor functioning, and emotional health. The NIH Toolbox for Assessment of Neurological and Behavioral Function[3] provides investigators with brief, yet comprehensive measurement tools for assessment of cognitive function, emotional health, sensory, and motor function. It provides an innovative approach to measurement that is responsive to the needs of researchers in a variety of settings, with a particular emphasis on measuring outcomes in clinical trials and functional status in large cohort studies (e.g., epidemiological studies and longitudinal studies). Included as subdomains of emotional health were negative affect, psychological well-being, stress and self-efficacy, and social relationships. Three PROMIS emotional distress item banks (Anger, Anxiety, and Depression) were used as measures of negative affect. Additionally, existing “legacy” measures, e.g., Patient Health Questionnaire (PHQ-9) and Center for Epidemiological Studies Depression Scale (CES-D), were flagged as potential candidates for the NIH Toolbox battery because of their history, visibility, and research legacy. Among these legacy measures, we focused on those that were available without proprietary restrictions for research applications. In most cases, these measures had been developed using classical test theory.

2.3. Quality of Life Outcomes in Neurological Disorders (Neuro-QoL)

The National Institute of Neurological Disorders and Stroke sponsored a multi-site project to develop clinically relevant and psychometrically robust Quality of Life (QOL) assessment tools for adults and children with neurological disorders. The primary goal of this effort, known as Neuro-QoL [4], was to enable clinical researchers to compare the QOL impact of different interventions within and across various conditions. This resulted in 13 adult QOL item banks (Anxiety, Depression, Fatigue, Upper Extremity Function - Fine Motor, Lower Extremity Function - Mobility, Applied Cognition - General Concerns, Applied Cognition - Executive Function, Emotional and Behavioral Dyscontrol, Positive Affect and Well-Being, Sleep Disturbance, Ability to Participate in Social Roles and Activities, Satisfaction with Social Roles and Activities, and Stigma), eight pediatric item banks (Anger, Anxiety, Depression, Fatigue, Pain, Applied Cognition - General Concerns, Social Relations - Interaction with Peers, and Stigma) and two additional pediatric physical function scales (Lower Extremity Function –Mobility, and Upper Extremity Function -Fine Motor, ADL).

[3] www.nihtoolbox.org
[4] www.neuroqol.org

3. Legacy Instruments

Typically, we have linked widely accepted “legacy” measures that were part of the initial validation work for PROMIS or NIH Toolbox. In some cases, instruments were administered as part of the PROsetta Stone project for specific linking purposes. In this case, we have linked Pediatric Perceived Cognitive Function to Neuro-QoL Pediatric Cognitive Function, where the former serves as the “legacy” instrument. Data were collected on reference measures (e.g., PROMIS Depression) from a minimum of 400 respondents (for stable item parameter estimation), along with responses to at least one other conceptually similar scale or bank to be linked to the reference measure. (See Table 5.1).

3.1. Pediatric Perceived Cognitive Function Item Bank (Peds PCF)

The Pediatric Perceived Cognitive Function Item Bank (Peds PCF) consists of 43 items measuring children’s cognitive behaviors. Both parent-reported and child-reported versions are available. The Peds PCF was initially designed for children with cancer who receive neurotoxicity treatments and for other populations of children and adolescents at risk for neurocognitive impairment. The Peds PCF has satisfactory psychometric properties, as evaluated using both classical test theory and IRT approaches, in use with the US general population and with children with cancer (Lai et al, 2011). It produces reliable scores that can discriminate between children with (versus without) significant symptoms of attention, social, and thought problems as well as between children with brain tumors versus those with other types of cancer. US general population-based norms are available to serve as a reference. This measure uses two 5-point rating scales: One is frequency related: (“none of the time” to “all of the time”) and one is intensity related (“not at all” to “very much”). For context, a 4-week timeframe is used. A 7-item short form and a computer adaptive test (CAT) version of the item bank are available. In 2015, the Peds PCF was submitted for inclusion in PROMIS as the PROMIS Pediatric Cognitive Function item bank. The PROMIS Pediatric Cognitive Function instrument is scored on the T-score metric, not raw scores as included in the linking tables within this document. Therefore, PROMIS Pediatric Cognitive Function should not utilize the linking tables provided here.

4. Linking Methods

PROMIS full-bank administration allows for single-group linking. This linking method is used when two or more measures are administered to the same group of people. For example, two PROMIS banks (Anxiety and Depression) and three legacy measures (MASQ, CES-D, and SF-36 MH) were administered to a sample of 925 people, with the order of measures presented randomized so as to minimize potential order effects. The original purpose of the PROMIS full-bank administration study was to establish initial validity evidence (e.g., validity coefficients), not to establish linking relationships. Thus, initial analyses of the full-bank administration sample revealed several potential score-linking issues: (a) some measures had severely skewed score distributions; (b) the sample size for some administered measures was relatively small. These score-linking issues can be limiting factors when determining an appropriate linking method (e.g., what method options are available or whether linking can even be conducted). Another potential linking issue is related to how the non-PROMIS measures are scored and reported. For example, all SF-36 subscales are scored using a proprietary scoring algorithm and reported as normed scores (0 to 100). Others are scored and reported using simple raw summed scores. All PROMIS measures are scored using the final re-centered item response theory (IRT) item parameters and transformed to the T-score metric (mean=50, SD=10).

PROMIS's T-score distributions are standardized so that a score of 50 represents the average (mean) for the US general population and the standard deviation around that mean is 10 points. A high PROMIS score always represents more of the concept being measured. Thus, a person who has a T-score of 60 is one standard deviation higher than the general population for the concept being measured. It therefore follows that, for condition symptoms and negatively-framed or oriented concepts like pain, fatigue, and anxiety, a score of 60 is one standard deviation worse than average; while for functional scores and positively-framed or oriented concepts like physical and social function, a score of 60 is one standard deviation better than average.

In order to apply linking methods consistently across different studies, linking/concordance relationships were established based on the raw summed score metric of the measures. Furthermore, the direction of linking relationships established was from legacy to PROMIS measure. That is, each raw summed score on a given legacy instrument was mapped to a T-score on the corresponding PROMIS instrument/bank. Finally, the raw summed score for each legacy instrument was constructed so that higher scores would represent higher levels of the construct being measured (to be consistent with the PROMIS approach). When legacy measures were scaled in the opposite direction, we reversed the direction of the legacy measure in order for the correlation between legacy and PROMIS measures to be positive and thereby facilitate concurrent calibration. As a result, some or all item response scores for some legacy instruments needed to be reverse-coded.

4.1. IRT Linking

One of the objectives of the current linking analyses is to determine whether the non-PROMIS measures can be added to their respective PROMIS item banks without significantly altering the underlying trait being measured. The rationale is twofold: (1) the augmented PROMIS item banks might provide more robust coverage, both in terms of content and difficulty; and (2) calibrating the non-PROMIS measures on the corresponding PROMIS item bank scale might facilitate subsequent linking analyses. At least two IRT linking approaches are applicable under the current study design: (1) linking separate calibrations through the Stocking-Lord method and (2) fixed-parameter calibration.

Linking separate calibrations might involve the following steps under the current setting.

- First, simultaneously calibrate the combined item set (e.g., PROMIS Depression bank and CES-D).
- Second, estimate linear transformation coefficients (additive and multiplicative constants) using the item parameters for the PROMIS bank items as anchor items.
- Third, transform the metric for the non-PROMIS items to the PROMIS metric.

The second approach, fixed-parameter calibration, involves fixing the PROMIS item parameters at their final bank values and calibrating only non-PROMIS items in order that the non-PROMIS item parameters may be placed on the same metric as the PROMIS items; that is, the focus is on placing the parameters of non-PROMIS items on the PROMIS metric. Updating the PROMIS item parameters is not desired, because the larger PROsetta-wide linking exercise is built on the stability of these final PROMIS calibrations. Note that IRT linking would be necessary when the ability level of the full-bank testing sample is different from that of the PROMIS scale-setting sample. If it is assumed that the two samples are from the same population, linking is not necessary and calibration of the items (either separately or simultaneously) will result in item parameter estimates that are on the same scale metric without any further scale linking. Even though the full-bank testing sample was a subset of the full PROMIS calibration sample, it is still possible that the two samples are somewhat disparate due to some non-random component of the selection process. Moreover, there is some evidence that linking can improve the accuracy of parameter estimation even when linking is not fully necessary (e.g., two samples are from the same population having the same or similar ability levels). Thus, conducting IRT linking would be worthwhile, with potential score accuracy benefits gained.

Once the non-PROMIS items are calibrated on the corresponding PROMIS item bank metric, the augmented item bank can be used for standard computation of IRT scaled scores from any subset of the items, including computerized adaptive testing (CAT) and creating short forms. The non-PROMIS items will be treated the same as the existing PROMIS items. Again, the above options are feasible only when the dimensionality of the bank is not altered significantly (i.e., where a unidimensional IRT model remains suitable for the aggregate set of items). Thus, prior to conducting IRT linking, it is important to assess the dimensionality of the involved measures based on separate and combined PROMIS and non-PROMIS measures. Various dimensionality assessment tools can be used, including confirmatory factor analysis, disattenuated correlations, and essential unidimensionality.

4.2. Equipercentile Linking

The IRT linking procedures described above are permissible only if the traits being measured are not significantly altered by aggregating items from multiple measures. One potential issue might be the creation of multidimensionality as a result of aggregating items measuring different traits. For two scales that measure distinct but highly related traits, predicting scores on one scale from those of the other has been used frequently. Concordance tables between PROMIS and non-PROMIS measures can be constructed using equipercentile equating (Lord, 1982; Kolen & Brennan, 2004) when there is insufficient empirical evidence that the instruments measure the same construct. An equipercentile method estimates a nonlinear linking relationship using percentile rank distributions of the two linking measures. The equipercentile linking method can be used in conjunction with a presmoothing method such as the loglinear model (Hanson, Zeng, & Colton, 1994). The frequency distributions are first smoothed using the loglinear model and then equipercentile linking is conducted based on the smoothed frequency distributions of the two measures. Smoothing can also be done at the backend on equipercentile equivalents and is called postsmoothing (Brennan, 2004; Kolen & Brennan, 2004). The cubic-spline smoothing algorithm (Reinsch, 1967) is used in the LEGS program employed in PROsetta analyses (Brennan, 2004). Smoothing is intended to reduce sampling error involved in the linking process. A successful linking procedure will provide a conversion (crosswalk) table, in which, for example, raw summed scores on the PHQ-9 measure are transformed to the T-score equivalents of the PROMIS Depression measure.

In the current context, equipercentile crosswalk tables can be generated using two different approaches. First is a direct linking approach where each raw summed score on a non-PROMIS measure is mapped directly to a PROMIS T-score. That is, raw summed scores on the non-PROMIS instrument and IRT scaled scores on the PROMIS (reference) instrument are linked directly, although raw summed scores and IRT scaled scores have distinct properties (e.g., discrete vs. continuous). This approach might be appropriate when the reference instrument is either an item bank or composed of a large number of items and so various subsets (static or dynamic) are likely to be used but not the full bank in its entirety (e.g., the PROMIS Physical Function bank with 124 items). Second is an indirect approach where raw summed scores on the non-PROMIS instrument are mapped to raw summed scores on the PROMIS instrument, and then the resulting raw summed score equivalents are mapped to corresponding scaled scores based on a raw-to-scale score conversion table. Because the raw summed score equivalents may take fractional values, such a conversion table will need to be interpolated using statistical procedures (e.g., cubic spline).

Finally, when samples are small or inadequate for a specific method, random sampling error becomes a major concern (Kolen & Brennan, 2004). That is, substantially different linking relationships might be obtained if linking is conducted repeatedly over different samples. This type of random sampling error can be measured by the standard error of equating (SEE), which can be operationalized as the standard deviation of equated scores for a given raw summed score over replications (Lord, 1982).

4.3. Linking Assumptions

In Section 5 of this PROsetta Stone report, we present the results of a large number of linking studies using a combination of newly collected and secondary data sets. In most cases, we have applied all three linking methods described in sections 4.1 and 4.2. Our purpose is to provide the maximum amount of useful information. However, the suitability of these methods depends upon the meeting of various linking assumptions. These assumptions require that the two instruments to be linked measure the same construct, show a high correlation, and are relatively invariant in subpopulation differences (Dorans, 2007). The degree to which these assumptions are met varies across linking studies. Given that different researchers may interpret these requirements differently, we have taken a liberal approach for inclusion of linkages in this book. Nevertheless, we recommend that researchers diagnostically review the classical psychometrics and CFA results in light of these assumptions prior to any application of the cross-walk charts or legacy parameters to their own data. Having investigated a large number of possible links between PROMIS measures and legacy measures, we did apply a few minimal exclusion rules before linking. For example, we generally did not proceed with planned linking when the raw score correlation between two instruments was less than .70.

5. Linking Results

Table 5.1 lists the linking analyses included in this report, which have been conducted based on samples from a Pediatric PCF study (see Section 2 for more details). In most cases, PROMIS instruments were used as the reference (i.e., scores on non-PROMIS instruments are expressed on the PROMIS score metric).

Table 5.1. Linking by Reference Instrument

Section	PROMIS Instrument	Instrument to Link	Study
5.8	Neuro-QoL v2.0 Pediatric Cognitive Function*	Pediatric Perceived Cognitive Function	PROsetta Stone Wave 3

** In 2014, the Neuro-QoL Pediatric Applied Cognition -- General Concerns bank was modified to be the v2.0 Pediatric Cognitive Function item bank.*

5.8 Neuro-QoL Pediatric Cognitive Function and Pediatric Perceived Cognitive Function (Peds PCF)

In this section we provide a summary of the procedures employed to establish a crosswalk between two measures of cognition, namely the Neuro-QoL v2.0 Pediatric Cognitive Function (NQ Peds Cog) item bank (14 items) and Peds PCF (30 items). Both instruments were scaled so that higher scores represent higher levels of cognition. We created raw summed scores for each of the measures separately and then for them combined. Summing of item scores assumes that all items have positive correlations with the total as examined in the section on Classical Item Analysis. Our sample consisted of 507 participants (N = 505 for participants with complete responses).

5.8.1 Raw Summed Score Distribution

The maximum possible raw summed scores were 70 for NQ Peds Cog and 150 for Peds PCF. Figure 5.8.1 and Figure 5.8.2 graphically display the raw summed score of the two measures. Figure 5.8.3 shows the distribution for them combined. Figure 5.8.4 is a scatter plot matrix showing the relationship of each pair of raw summed scores. Pearson correlations are shown above the diagonal. The correlation between NQ Peds Cog and Peds PCF was 0.93. The disattenuated (corrected for unreliabilies) correlation between NQ Peds Cog and Peds PCF was 0.96. The correlations between the combined score and the measures were
0.97 and 0.99 for NQ Peds Cog and Peds PCF, respectively.

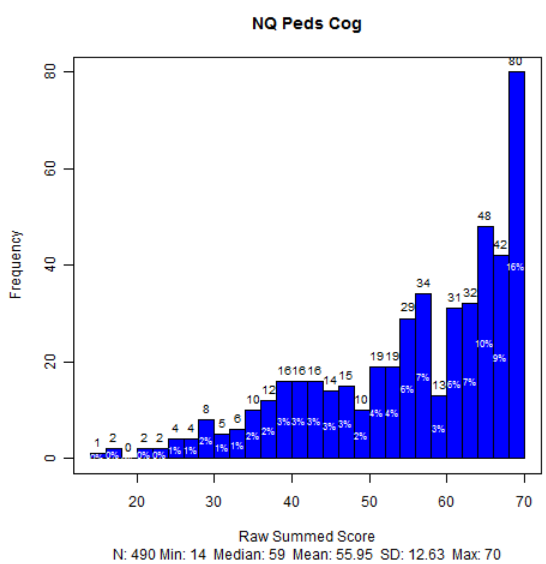

Figure 5.8.1: Raw Summed Score Distribution - Neuro-QoL Peds Cognitive Function

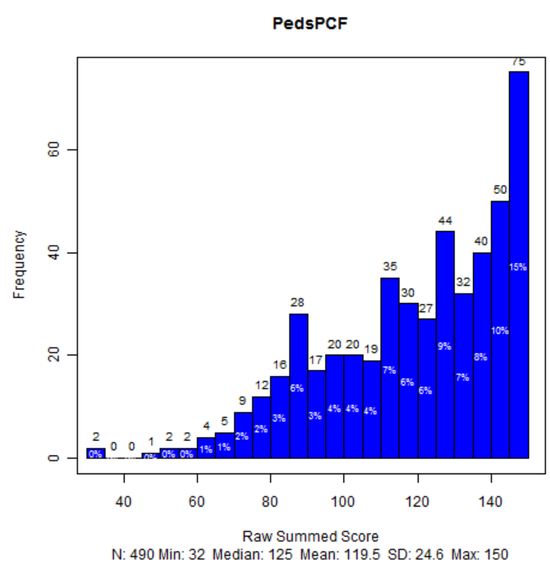

Figure 5.8.2: Raw Summed Score Distribution – Peds PCF

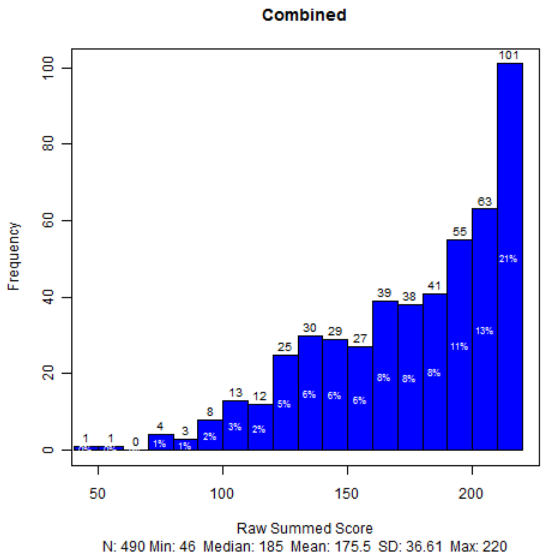

Figure 5.8.3: Raw Summed Score Distribution – Combined

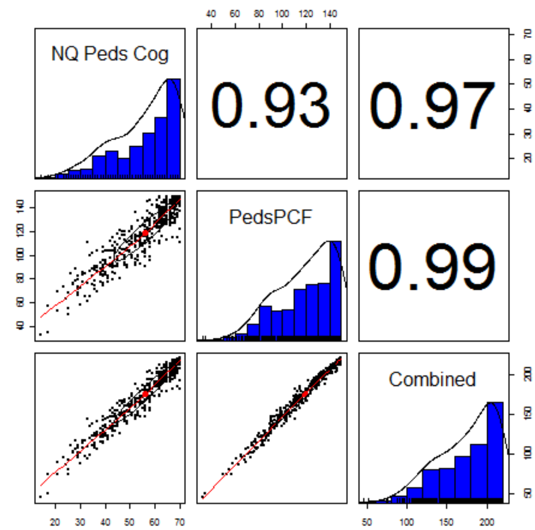

Figure 5.8.4: Scatter Plot Matrix of Raw Summed Scores

5.8.2 Classical Item Analysis

We conducted classical item analyses on the two measures separately and on them combined. Table 5.8.1 summarizes the results. For NQ Peds Cog, Cronbach's alpha internal consistency reliability estimate was 0.959 and adjusted (corrected for overlap) item-total correlations ranged from 0.712 to 0.833. For Peds PCF, alpha was 0.975 and adjusted item-total correlations ranged from 0.535 to 0.826. For the 44 items, alpha was 0.983 and adjusted item-total correlations ranged from 0.52 to 0.837.

Table 5.8.1: Classical Item Analysis

	No. Items	Cronbach's Alpha Internal Consistency Reliability Estimate	Adjusted (corrected for overlap) Item-total Correlation		
			Minimum	Mean	Maximum
NQ Peds Cog	14	0.959	0.712	0.774	0.833
Peds PCF	30	0.975	0.535	0.740	0.826
Combined	44	0.983	0.520	0.749	0.837

5.8.3 Confirmatory Factor Analysis (CFA)

To assess the dimensionality of the measures, a categorical confirmatory factor analysis (CFA) was carried out using the WLSMV estimator of Mplus on a subset of cases without missing responses. A single factor model (based on polychoric correlations) was run on each of the two measures separately and on the combined. Table 5.8.2 summarizes the model fit statistics. For NQ Peds Cog, the fit statistics were as follows: CFI = 0.977, TLI= 0.972, and RMSEA = 0.116.

For Peds PCF, CFI = 0.963, TLI = 0.960, and RMSEA = 0.081. For the 44 items, CFI = 0.951, TLI = 0.949, and RMSEA = 0.81.The main interest of the current analysis is whether the combined measure is essentially unidimensional.

Table 5.8.2: CFA Fit Statistics

	No. Items	n	CFI	TLI	RMSEA
NQ Peds Cog	14	507	0.977	0.972	0.116
Peds PCF	30	507	0.963	0.960	0.081
Combined	44	507	0.951	0.949	0.081

5.8.4 Item Response Theory (IRT) Linking

We conducted concurrent calibration on the combined set of 44 items according to the graded response model. The calibration was run using MULTILOG and two different approaches as described previously (i.e., IRT linking vs. fixed-parameter calibration). For IRT linking, all 44 items were calibrated freely on the conventional theta metric (mean=0, SD=1). Then the 14 NQ Peds Cog items served as anchor items to transform the item parameter estimates for the Peds PCF items onto the NQ Peds Cog metric. We used four IRT linking methods implemented in plink (Weeks, 2010): mean/mean, mean/sigma, Haebara, and Stocking-Lord. The first two methods are based on the mean and standard deviation of item parameter estimates, whereas the latter two are based on the item and test information curves. Table 5.8.3 shows the additive (A) and multiplicative (B) transformation constants derived from the four linking methods. For fixed-parameter calibration, the item parameters for the NQ Peds Cog items were constrained to their final bank values, while the Peds PCF items were calibrated, under the constraints imposed by the anchor items.

Table 5.8.3: IRT Linking Constants

	A	B
Mean/Mean	1.010	-0.322
Mean/Sigma	1.060	-0.290
Haebara	1.054	-0.294
Stocking-Lord	1.048	-0.298

The item parameter estimates for the Peds PCF items were linked to the NQ Peds Cog metric using the transformation constants shown in Table 5.8.3. The Peds PCF item parameter estimates from the fixed-parameter calibration are considered already on the NQ Peds Cog metric. Based on the transformed and fixed-parameter estimates we derived test characteristic curves (TCC) for Peds PCF as shown in Figure 5.8.5. Using the fixed-parameter calibration as a basis we then examined the difference with each of the TCCs from the four linking methods. Figure 5.8.6 displays the differences on the vertical axis.

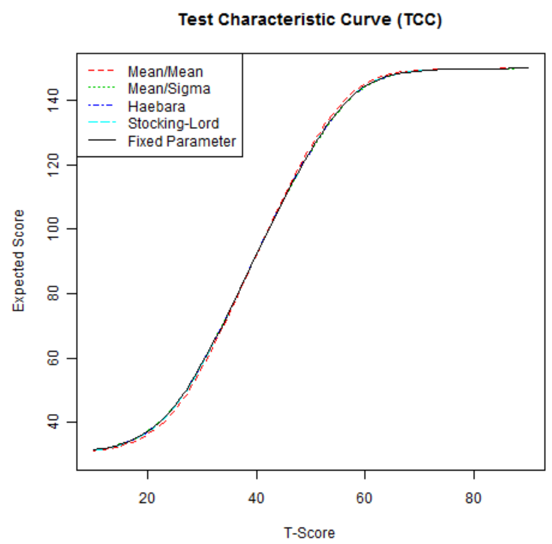

Figure 5.8.5: Test Characteristic Curves (TCC) from Different Linking Methods

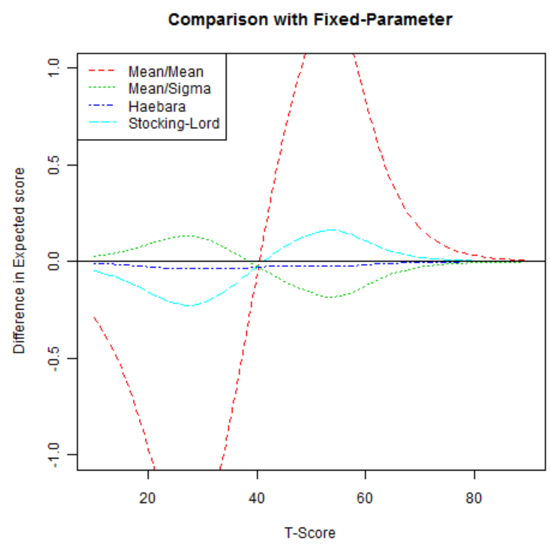

Figure 5.8.6: Difference in Test Characteristic Curves (TCC)

Table 5.8.4 shows the fixed-parameter calibration item parameter estimates for Peds PCF. The marginal reliability estimate for Peds PCF based on the item parameter estimates was 0.96. The marginal reliability estimates for NQ Peds Cog and the combined set were 0.926 and 0.971, respectively. The slope parameter estimates for Peds PCF ranged from 1.37 to 3.61 with a mean of 2.51. The slope parameter estimates for NQ Peds Cog ranged from 2.18 to 3.74 with a mean of 2.89. We also derived scale information functions based on the fixed-parameter calibration result. Figure 5.8.7 displays the scale information functions for NQ Peds Cog, Peds PCF, and the combined set of 44. We then computed IRT scaled scores for the three measures based on the fixed-parameter calibration result. Figure 5.8.8 is a scatter plot matrix showing the relationships between the measures.

Table 5.8.4: Fixed-Parameter Calibration Item Parameter Estimates for Peds PCF

a	cb1	cb2	cb3	cb4	NCAT
1.366	-3.753	-2.569	-1.406	-0.270	5
2.056	-2.399	-1.528	-0.466	0.539	5
1.827	-2.945	-1.849	-0.831	0.304	5
2.486	-2.005	-1.352	-0.507	0.418	5
2.648	-2.289	-1.547	-0.721	0.259	5
2.182	-2.622	-1.728	-1.008	-0.208	5
2.781	-2.220	-1.355	-0.629	0.255	5
1.754	-1.661	-0.798	0.115	1.167	5
2.345	-2.468	-1.722	-0.916	0.099	5
2.712	-2.287	-1.402	-0.566	0.330	5
2.427	-2.235	-1.340	-0.532	0.467	5
2.328	-2.293	-1.493	-0.511	0.608	5
2.196	-2.607	-1.612	-0.733	0.329	5
2.142	-2.916	-1.971	-1.104	-0.312	5
2.159	-2.147	-1.466	-0.837	0.040	5
1.681	-3.109	-2.253	-1.172	-0.207	5
2.148	-2.176	-1.269	-0.391	0.719	5
2.805	-1.902	-1.184	-0.453	0.474	5
2.488	-2.305	-1.355	-0.531	0.491	5
2.496	-2.515	-1.797	-0.806	0.242	5
2.734	-2.219	-1.317	-0.545	0.451	5
2.548	-2.567	-1.592	-0.906	-0.202	5
2.796	-2.349	-1.452	-0.625	0.247	5
3.050	-1.828	-1.244	-0.555	0.206	5
3.218	-2.011	-1.480	-0.716	0.100	5
2.409	-2.386	-1.597	-0.939	-0.190	5
3.238	-2.001	-1.394	-0.649	0.257	5
3.314	-1.861	-1.124	-0.453	0.466	5
3.609	-1.822	-1.153	-0.431	0.420	5
3.480	-1.939	-1.145	-0.536	0.311	5

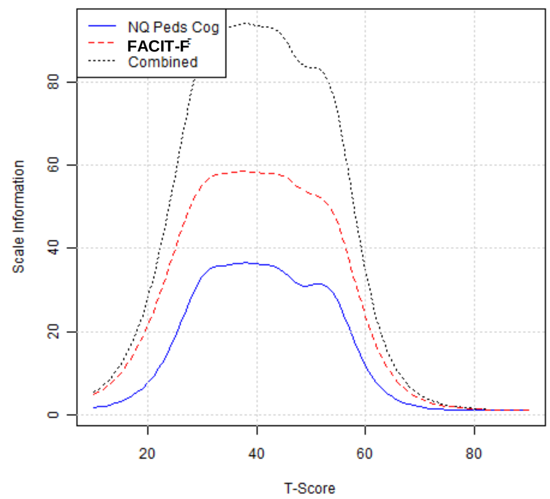

Figure 5.8.7: Comparison of Scale Information Functions

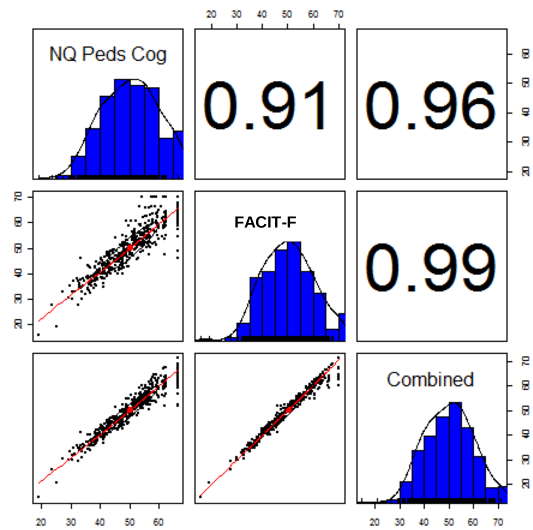

Figure 5.8.8: Comparison of IRT Scaled Scores

5.8.5 Raw Score to T-Score Conversion using Linked IRT Parameters

The IRT model implemented in PROMIS (i.e., the graded response model) uses the pattern of item responses for scoring, not just the sum of individual item scores. However, a crosswalk table mapping each raw summed score point on Peds PCF to a scaled score on NQ Peds Cog can be useful. Based on the Peds PCF item parameters derived from the fixed-parameter calibration, we constructed a score conversion table. The conversion table displayed in Appendix Table 21 can be used to map simple raw summed scores from Peds PCF to T-score values linked to the NQ Peds Cog metric. Each raw summed score point and corresponding NQ Peds Cog scaled score are presented along with the standard error associated with the scaled score. The raw summed score is constructed such that for each item, consecutive integers in base 1 are assigned to the ordered response categories.

5.8.6 Equipercentile Linking

We mapped each raw summed score point on Peds PCF to a corresponding scaled score on NQ Peds Cog by identifying scores on NQ Peds Cog that have the same percentile ranks as scores on Peds PCF. Theoretically, the equipercentile linking function is symmetrical for continuous random variables (X and Y). Therefore, the linking function for the values in X to those in Y is the same as that for the values in Y to those in X. However, for discrete variables like raw summed scores the equipercentile linking functions can be slightly different (due to rounding errors and differences in score ranges) and hence may need to be obtained separately. Figure 5.8.9 displays the cumulative distribution functions of the measures. Figure 5.8.10 shows the equipercentile linking functions based on raw summed scores, from Peds PCF to NQ Peds Cog. When the number of raw summed score points differs substantially, the equipercentile linking functions could deviate from each other noticeably. The problem can be exacerbated when the sample size is small. Appendix Table 22 and Appendix Table 23 show the equipercentile crosswalk tables. The result shown in Appendix Table 22 is based on the direct (raw summed score to scaled score) approach, whereas Appendix Table 23 shows the result based on the indirect (raw summed score to raw summed score equivalent to scaled score equivalent) approach (Refer to Section 4.2 for details). Three separate equipercentile equivalents are presented: one is equipercentile without post smoothing ("Equipercentile Scale Score Equivalents") and two with different levels of postsmoothing, i.e., "Equipercentile Equivalents with Postsmoothing (Less Smoothing)" and "Equipercentile Equivalents with Postsmoothing (More Smoothing)." Postsmoothing values of 0.3 and 1.0 were used for "Less" and "More," respectively (Refer to Brennan, 2004 for details).

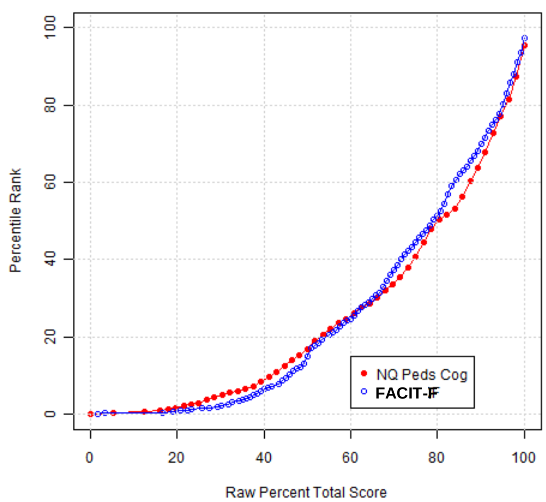

Figure 5.8.9: Comparison of Cumulative Distribution Functions based on Raw Summed Scores

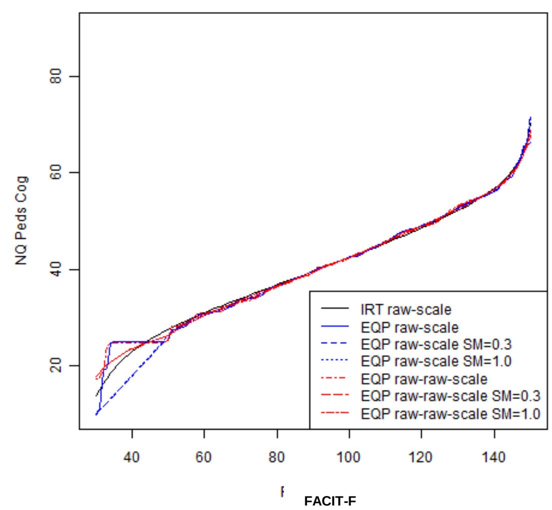

Figure 5.8.10: Equipercentile Linking Functions

5.8.7 Summary and Discussion

The purpose of linking is to establish the relationship between scores on two measures of closely related traits. The relationship can vary across linking methods and samples employed. In equipercentile linking, the relationship is determined based on the distributions of scores in a given sample. Although IRT-based linking can potentially offer sample-invariant results, they are based on estimates of item parameters, and hence subject to sampling errors. A potential issue with IRT-based linking methods is, however, the violation of model assumptions as a result of combining items from two measures (e.g., unidimensionality and local independence). As displayed in Figure 5.8.10, the relationships derived from various linking methods are consistent, which suggests that a robust linking relationship can be determined based on the given sample.

To further facilitate the comparison of the linking methods, Table 5.8.5 reports four statistics summarizing the current sample in terms of the differences between the NQ Peds Cog T-scores and Peds PCF scores linked to the T-score metric through different methods. In addition to the seven linking methods previously discussed (see Figure 5.8.10), the method labeled “IRT pattern scoring” refers to IRT scoring based on the pattern of item responses instead of raw summed scores. With respect to the correlation between observed and linked T-scores, IRT pattern scoring produced the best result (0.911), followed by EQP raw-raw-scale SM=0.3 (0.91). Similar results were found in terms of the standard deviation of differences and root mean squared difference (RMSD). EQP raw-scale SM=0.0 yielded smallest RMSD (4.067), followed by EQP raw-raw- scale SM=0.3 (4.071).

Table 5.8.5: Observed vs. Linked T-scores

Methods	Correlation	Mean Difference	SD Difference	RMSD
IRT pattern scoring	0.911	-0.079	4.143	4.140
IRT raw-scale	0.903	-0.176	4.320	4.320
EQP raw-scale SM=0.0	0.909	0.009	4.071	4.067
EQP raw-scale SM=0.3	0.901	-0.245	4.424	4.426
EQP raw-scale SM=1.0	0.902	-0.266	4.379	4.382
EQP raw-raw-scale SM=0.0	0.909	-0.052	4.091	4.087
EQP raw-raw-scale SM=0.3	0.910	0.015	4.075	4.071
EQP raw-raw-scale SM=1.0	0.907	-0.029	4.172	4.168

To examine the bias and standard error of the linking results, a resampling study was used. In this procedure, small subsets of cases (e.g., 25, 50, and 75) were drawn with replacement from the study sample (N=490) over a large number of replications (i.e., 10,000).

Table 5.8.6 summarizes the mean and standard deviation of differences between the observed and linked T-scores by linking method and sample size. For each replication, the mean difference between the observed and equated NQ Peds Cog T-scores was computed. Then the mean and the standard deviation of the means were computed over replications as bias and empirical standard error, respectively. As the sample size increased (from 25 to 75), the empirical standard error decreased steadily. At a sample size of 75, EQP raw-raw-scale SM=0.0 produced the smallest standard error, 0.432. That is, the difference between the mean NQ Peds Cog T-score and the mean equated Peds PCF T-score based on a similar sample of 75 cases is expected to be around ±0.86 (i.e., 2 × 0.432).

Table 5.8.6: Comparison of Resampling Results

Methods	Mean (N=25)	SD (N=25)	Mean (N=50)	SD (N=50)	Mean (N=75)	SD (N=75)
IRT pattern scoring	-0.083	0.804	-0.084	0.548	-0.069	0.438
IRT raw-scale	-0.172	0.844	-0.171	0.580	-0.179	0.461
EQP raw-scale SM=0.0	0.010	0.797	0.002	0.544	0.009	0.435
EQP raw-scale SM=0.3	-0.233	0.863	-0.248	0.596	-0.237	0.468
EQP raw-scale SM=1.0	-0.277	0.854	-0.266	0.586	-0.268	0.465
EQP raw-raw-scale SM=0.0	-0.058	0.794	-0.059	0.546	-0.056	0.432
EQP raw-raw-scale SM=0.3	0.007	0.788	0.019	0.545	0.012	0.438
EQP raw-raw-scale SM=1.0	-0.030	0.813	-0.034	0.561	-0.028	0.447

Examining a number of linking studies in the current project revealed that the two linking methods (IRT and equipercentile) in general produced highly comparable results. Some noticeable discrepancies were observed (albeit rarely) in some extreme score levels where data were sparse. Model-based approaches can provide more robust results than those relying solely on data when data are sparse. The caveat is that the model should fit the data reasonably well. One of the potential advantages of IRT-based linking is that the item parameters on the linking instrument can be expressed on the metric of the reference instrument, and therefore can be combined without significantly altering the underlying trait being measured. As a result, a larger item pool might be available for computerized adaptive testing or various subsets of items can be used in static short forms. Therefore, IRT-based linking (Appendix Table 21) might be preferred when the results are comparable and no apparent violations of assumptions are evident.

6. Appendix Table 21: Raw Score to T-Score Conversion Table (IRT Fixed Parameter Calibration Linking) for Pediatric PCF 30-Item Short Form and Neuro-QoL Pediatric Cognitive Function (PROsetta Stone Wave 3 Study) - RECOMMENDED

Peds PCF Raw Score	Neuro-QoL Ped Cognitive Function T-Score	SE	Peds PCF Raw Score	Neuro-QoL Ped Cognitive Function T-Score	SE
30	13.9	2.4	91	39.9	1.4
31	15.0	2.6	92	40.1	1.4
32	16.2	2.7	93	40.4	1.4
33	17.3	2.6	94	40.7	1.4
34	18.4	2.6	95	41.0	1.4
35	19.3	2.4	96	41.3	1.4
36	20.2	2.3	97	41.5	1.4
37	20.9	2.2	98	41.8	1.4
38	21.7	2.1	99	42.1	1.4
39	22.3	2.0	100	42.4	1.4
40	22.9	1.9	101	42.7	1.4
41	23.5	1.9	102	43.0	1.4
42	24.0	1.8	103	43.3	1.4
43	24.5	1.8	104	43.6	1.4
44	25.0	1.7	105	43.9	1.4
45	25.5	1.7	106	44.2	1.4
46	25.9	1.7	107	44.4	1.4
47	26.3	1.6	108	44.7	1.4
48	26.7	1.6	109	45.0	1.4
49	27.1	1.6	110	45.4	1.4
50	27.5	1.6	111	45.7	1.4
51	27.9	1.5	112	46.0	1.4
52	28.3	1.5	113	46.3	1.4
53	28.6	1.5	114	46.6	1.4
54	29.0	1.5	115	46.9	1.4
55	29.3	1.5	116	47.2	1.4
56	29.7	1.5	117	47.5	1.4
57	30.0	1.5	118	47.9	1.4
58	30.3	1.4	119	48.2	1.4
59	30.6	1.4	120	48.5	1.4
60	30.9	1.4	121	48.9	1.5
61	31.3	1.4	122	49.2	1.5
62	31.6	1.4	123	49.6	1.5
63	31.9	1.4	124	49.9	1.5
64	32.2	1.4	125	50.3	1.5
65	32.5	1.4	126	50.6	1.5

66	32.8	1.4	127	51.0	1.5
67	33.1	1.4	128	51.4	1.5
68	33.4	1.4	129	51.7	1.5
69	33.7	1.4	130	52.1	1.5
70	33.9	1.4	131	52.5	1.5
71	34.2	1.4	132	52.9	1.5
72	34.5	1.4	133	53.3	1.5
73	34.8	1.4	134	53.8	1.6
74	35.1	1.4	135	54.2	1.6
75	35.4	1.4	136	54.7	1.6
76	35.7	1.4	137	55.2	1.7
77	35.9	1.4	138	55.7	1.7
78	36.2	1.4	139	56.2	1.7
79	36.5	1.4	140	56.8	1.8
80	36.8	1.4	141	57.4	1.9
81	37.1	1.4	142	58.0	2.0
82	37.3	1.4	143	58.7	2.1
83	37.6	1.4	144	59.6	2.2
84	37.9	1.4	145	60.5	2.4
85	38.2	1.4	146	61.6	2.6
86	38.5	1.4	147	62.8	2.9
87	38.7	1.4	148	64.4	3.3
88	39.0	1.4	149	66.6	3.7
89	39.3	1.4	150	70.3	4.9
90	39.6	1.4			

7. Appendix Table 22: Direct (Raw to Scale) Equipercentile Crosswalk Table – From Pediatric PCF 30-Item Short Form to Neuro-QoL Pediatric Cognitive Function – Table 21 is recommended

Peds PCF Raw Score	Equipercentile PROMIS Scaled Score Equivalents (No Smoothing)	Equipercentile Equivalents with Postsmoothing (Less Smoothing)	Equipercentile Equivalents with Postsmoothing (More Smoothing)	Standard Error of Equating (SEE)
30	10	10	10	0.35
31	11	11	11	0.35
32	19	11	12	0.35
33	20	12	12	0.35
34	25	13	13	0.35
35	25	14	14	0.35
36	25	15	15	0.35
37	25	15	16	0.35
38	25	16	16	0.35
39	25	17	17	0.35
40	25	18	18	0.35
41	25	19	19	0.35
42	25	19	20	0.35
43	25	20	20	0.35
44	25	21	21	0.35
45	25	22	22	0.35
46	25	23	23	0.35
47	25	23	24	0.35
48	25	24	25	0.35
49	25	25	25	0.35
50	25	26	26	0.35
51	28	27	27	0.71
52	28	27	27	0.71
53	28	28	28	0.71
54	28	28	28	0.79
55	28	29	29	0.79
56	30	29	29	2.00
57	30	30	29	2.00
58	31	30	30	0.50
59	31	30	30	0.52
60	31	31	30	0.52
61	31	31	31	0.52
62	31	31	31	0.56
63	31	31	31	0.56
64	31	31	31	0.54
65	32	32	32	0.77

66	32	32	32	0.90
67	32	32	32	1.02
68	32	33	33	1.70
69	34	33	33	0.34
70	34	33	33	0.32
71	34	34	34	0.32
72	34	34	34	0.34
73	34	34	34	0.34
74	34	34	34	0.32
75	34	35	35	0.35
76	35	35	35	2.15
77	36	36	36	0.71
78	36	36	36	0.70
79	36	36	36	0.45
80	37	37	37	0.43
81	37	37	37	0.44
82	37	37	37	0.44
83	37	37	37	0.46
84	38	38	38	0.41
85	38	38	38	0.40
86	38	38	38	0.43
87	39	39	39	0.32
88	39	39	39	0.31
89	39	39	39	0.27
90	40	40	40	0.33
91	40	40	40	0.34
92	40	40	40	0.35
93	41	41	41	0.35
94	41	41	41	0.35
95	41	41	41	0.35
96	41	41	41	0.35
97	42	42	42	0.43
98	42	42	42	0.43
99	42	42	42	0.43
100	42	42	42	0.45
101	43	43	43	0.44
102	43	43	43	0.44
103	43	43	43	0.42
104	43	44	44	0.43
105	44	44	44	0.64
106	44	44	44	0.62
107	44	45	45	0.61
108	45	45	45	0.69
109	45	45	45	0.72
110	46	46	46	0.41
111	46	46	46	0.41
112	47	47	47	0.54
113	47	47	47	0.52

114	48	47	47	0.24
115	48	48	48	0.23
116	48	48	48	0.22
117	48	48	48	0.22
118	48	48	48	0.22
119	49	49	49	0.38
120	49	49	49	0.36
121	49	49	49	0.37
122	50	50	50	0.28
123	50	50	50	0.28
124	50	50	50	0.27
125	50	51	51	0.27
126	51	51	51	0.92
127	52	51	51	0.23
128	52	52	52	0.22
129	52	52	52	0.23
130	53	53	53	0.48
131	53	53	53	0.46
132	54	53	53	0.48
133	54	54	54	0.48
134	54	54	54	0.48
135	55	54	54	0.31
136	55	55	55	0.30
137	55	55	55	0.30
138	55	55	56	0.30
139	56	56	56	0.37
140	56	56	57	0.33
141	56	57	57	0.33
142	58	57	58	0.36
143	58	58	58	0.35
144	59	59	59	0.35
145	59	60	60	0.31
146	62	61	61	0.20
147	62	62	63	0.18
148	66	65	64	0.16
149	66	66	65	0.12
150	66	72	71	0.11

8. Appendix Table 23: Indirect (Raw to Raw to Scale) Equipercentile Crosswalk Table – From Pediatric PCF 30-Item Short Form to Neuro-QoL Pediatric Cognitive Function – Table 21 is recommended

Peds PCF Raw Score	Equipercentile PROMIS Scaled Score Equivalents (No Smoothing)	Equipercentile Equivalents with Postsmoothing (Less Smoothing)	Equipercentile Equivalents with Postsmoothing (More Smoothing)
30	17	18	18
31	17	19	19
32	19	19	19
33	24	20	20
34	24	21	21
35	25	21	21
36	25	22	22
37	25	22	22
38	25	23	23
39	25	23	23
40	25	23	23
41	25	24	24
42	25	24	24
43	25	24	24
44	25	25	25
45	25	25	25
46	25	25	25
47	25	25	26
48	25	26	26
49	25	26	26
50	25	26	26
51	28	27	27
52	28	28	27
53	28	28	28
54	28	29	28
55	29	29	28
56	30	29	29
57	30	30	29
58	31	30	30
59	31	30	30
60	31	31	30
61	31	31	31
62	31	31	31
63	31	31	31
64	31	32	32
65	32	32	32
66	32	32	32
67	32	32	33
68	33	33	33

69	33	33	33
70	33	33	34
71	33	34	34
72	34	34	34
73	34	34	34
74	34	34	35
75	34	35	35
76	35	35	35
77	36	35	36
78	36	36	36
79	36	36	36
80	37	36	36
81	37	37	37
82	37	37	37
83	37	37	37
84	38	38	38
85	38	38	38
86	38	38	38
87	38	38	38
88	38	39	39
89	39	39	39
90	40	39	39
91	40	40	40
92	40	40	40
93	41	40	40
94	41	41	41
95	41	41	41
96	41	41	41
97	42	42	42
98	42	42	42
99	42	42	42
100	42	42	42
101	43	43	43
102	43	43	43
103	43	43	43
104	44	44	44
105	44	44	44
106	44	44	44
107	45	45	45
108	45	45	45
109	45	45	45
110	45	46	46
111	46	46	46
112	47	46	46
113	47	47	46
114	47	47	47
115	48	48	47
116	48	48	48
117	48	48	48

118	48	48	48
119	48	48	48
120	49	49	49
121	49	49	49
122	49	49	50
123	49	50	50
124	50	50	50
125	50	50	50
126	51	51	51
127	51	51	51
128	52	52	52
129	53	52	52
130	53	53	52
131	54	53	53
132	54	54	53
133	54	54	54
134	54	54	54
135	55	55	54
136	55	55	55
137	55	55	55
138	56	56	56
139	56	56	56
140	57	57	57
141	57	57	57
142	58	58	58
143	58	58	58
144	59	59	59
145	60	60	60
146	61	61	61
147	62	62	62
148	64	63	64
149	65	65	66
150	68	68	69

9. Pediatric Perceived Cognitive Function 30-Item Short Form

ItemID	Item Context	Item Stem	Responses (Higher Scores = Better Function)
CaPSPC10_2C	In the past 4 weeks,	It is hard for me to find my way to a place that I have visited several times before	5= None of the time 4= A little of the time 3= Some of the time 3= Most of the time 1= All of the time
CaPSPC11_2C	In the past 4 weeks,	I have trouble remembering where I put things, like my watch or my homework	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
CaPSPC26_2C	In the past 4 weeks,	I have trouble remembering the names of people I have just met	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
NQCOG40_2_1C	In the past 4 weeks,	It is hard for me to take notes in class	5= Not at all 4= A little bit 3= Somewhat 2= Quite a bit 1= Very much x= Not applicable (I have not been in school at all in the last 4 weeks)
NQCOG40_2C	In the past 4 weeks,	It is hard for me to learn new things	5= Not at all 4= A little bit 3= Somewhat 2= Quite a bit 1= Very much
NQCOG42_2C	In the past 4 weeks,	It is hard for me to understand pictures that show how to make something	5= Not at all 4= A little bit 3= Somewhat 2= Quite a bit 1= Very much

NQCOG44_2_2C	In the past 4 weeks,	It is hard for me to pay attention to something boring I have to do	5= Not at all 4= A little bit 3= Somewhat 2= Quite a bit 1= Very much
NQCOG44_2C	In the past 4 weeks,	It is hard for me to pay attention to one thing for more than 5-10 minutes	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
NQCOG62_2C	In the past 4 weeks,	I have trouble recalling the names of things	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
NQCOG65_2C	In the past 4 weeks,	I have trouble keeping track of what I am doing if I get interrupted	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
NQCOG66_2C	In the past 4 weeks,	It is hard for me to do more than one thing at a time	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
NQCOG68_2_1C	In the past 4 weeks,	I forget what my parents or teachers ask me to do	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
NQCOG69_2C	In the past 4 weeks,	I walk into a room and forget what I wanted to get or do	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time

NQCOG70_2C	In the past 4 weeks,	I have trouble remembering the names of people I know	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
NQCOG78_2_1C	In the past 4 weeks,	It is hard for me to add or subtract numbers in my head	5= Not at all 4= A little bit 3= Somewhat 2= Quite a bit 1= Very much
NQCOG82_2C	In the past 4 weeks,	I have trouble remembering the date or day of the week	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
NQCOG85_2C	In the past 4 weeks,	When I have a big project to do, I have trouble deciding where to start	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
pB8_FC	In the past 4 weeks,	I forget things easily	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
pB9_FC	In the past 4 weeks,	It is hard for me to concentrate in school	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time x= Not applicable (I have not been in school at all in the last 4 weeks)
pB10_FC	In the past 4 weeks,	I have to read things several times to understand them	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time

pB11r_FC	In the past 4 weeks,	I react slower than most people my age when I play games	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
pB12_FC	In the past 4 weeks,	It is hard for me to find the right words to say what I mean	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
pedsPCF1_FC	In the past 4 weeks,	I have a hard time keeping track of my homework	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time x= Not applicable (I have not been in school at all in the last 4 weeks)
pedsPCF2_FC	In the past 4 weeks,	I forget to bring things to and from school that I need for homework	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time x= Not applicable (I have not been in school at all in the last 4 weeks)
pedsPCF3_FC	In the past 4 weeks,	I forget what I am going to say	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
pedsPCF4_FC	In the past 4 weeks,	It takes me longer than other people to get my school work done	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time x= Not applicable (I have not been in school at all in the last 4 weeks)

pedsPCF5_FC	In the past 4 weeks,	I have to use written lists more often than other people my age so I will not forget things	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
pedsPCF6_FC	In the past 4 weeks,	I have trouble remembering to do things like school projects or chores	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time
pedsPCF7_FC	In the past 4 weeks,	I have trouble paying attention to the teacher	5= None of the time 4= A little of the time 3= Some of the time 2= Most of the time 1= All of the time x= Not applicable (I have not been in school at all in the last 4 weeks)
pedsPCF8_FC	In the past 4 weeks,	I have to work really hard to pay attention or I make mistakes	1= None of the time 2= A little of the time 3= Some of the time 4= Most of the time 5= All of the time