XIX IMEKO World Congress
Fundamental and Applied Metrology
September 6-11, 2009, Lisbon, Portugal

FUNDAMENTAL MEASUREMENT OF FUNCTIONAL CAREGIVING IN REHABILIATION MEDICINE

Nikolaus Bezruczko [1], *Shu-Pi C. Chen* [2], *Constance Hill* [3], *Joyce M. Chesniak* [3]

[1] Measurement and Evaluation Consulting, Chicago, USA, nbezruczko@msn.com
[2] Saint Xavier University, Chicago, USA, schen@uic.edu
[3] Children's Memorial Hospital, Chicago, USA, chill@childrensmemorial.org; jchesniak@childrensmemorial.org

Abstract - The purpose of this research was to develop an objective, linear measure of mothers' confidence to care for children assisted with medical technology in their homes. Three technologies were addressed, namely, trachesotomy, tracheostomy and ventilator, and BiPAP/CPAP. All mothers (N=53) were from Chicago, USA and primary caregivers for children assisted with medical technologies. The theoretical construct for this research is Functional Caregiving (FC) and survey methods were guided by a caregiver content matrix validated by content and clinical reviews. Field testing with target mothers provided initial raw scores, which were linearized with a Rasch model for rating scales. Raw scores were first analyzed with Principal Components Analysis, and then *Tracheostomy* item and mothers' item responses were transformed to an objective, equal-interval scale with a Rasch model. Measurement model results show sample separation for the *Tracheostomy* caregiving scale was 2.66 and reliability, .88. Model fit results were adequate and residual analysis did not identify major dimensionality threats. Although sample size was small, transformation of raw scores to linear scale appears very promising. Further research should attempt to replicate these results with larger samples, as well as cross-validate FC with a competency construct.

Keywords: Rasch model, linear measurement, fundamental measurement, caregiving

1. INTRODUCTION

Mothers with children assisted with medical technologies (CAMT) from Children's Memorial Hospital (CMH) in Chicago, USA responded to questionnaire items about their confidence to provide caregiving in their homes (Bezruczko, Chen, Hill, and Chesniak, 2008; see also Hill, et al., 2008). CMH is aware of the enormous burden mothers assume by caring for CAMT, and published literature has reported higher health and well-being risks for these mothers (see Thyen, Kuhlthau, and Perrin, 1999). To better inform hospitals about mothers' caregiving confidence, self-reported perceptions were first analyzed for conformity to a mathematical measurement model, and then measured on a linear scale.

1.1 Problem

Many conditions and problems in physical and rehabilitation medicine have psycho-social origins. While physics has made sweeping advances establishing hypothetical constructs in the physical universe, the domination of correlation-based statistics in social science has severely retarded development of fundamental measurement in the social universe. Virtually all contemporary measurement in social research is only conducted at the level of ordinal measurement. Consequently, with the exception of Rasch models, social sciences do not have objective, linear scales to measure psycho-social constructs. In this research, we present an example of a solution to this problem by applying a Rasch measurement model to the problem of measuring mothers' confidence to provide care for CAMT.

2. METHOD

2.2 Sample

All biological mothers with children meeting the following criteria were eligible to participate in this research.

- between ages birth and 18 years
- served by Pulmonary Habilitation Program, Otolaryngology, Pulmonary Services and 9 West Inpatient Unit at Children's Memorial Hospital, Chicago
- assisted by medical technology (tracheostomy, tracheostomy/ventilator, BiPAP/CPAP) to maintain oxygenation and ventilation

In addition, all mothers had received CAMT caregiver training at CMH.

2.2 Data

Instrument development occurred in three major phases, a) construct development including operational definition with rating scale items, b) content and clinical validation of item content, and c) statistical analysis of mothers' responses. The construct development phase required a comprehensive description of tasks and responsibilities required of mothers caring for CAMT in their homes. Consequently, an item pool was generated to define

operationally a caregiving construct, and these items were submitted to expert content and staff reviews. Finally, items were selected for survey forms. The overall instrument development plan follows below:

- Construct item content specification based on caregiving literature review (see Fleming, 2004)
- Caregiving item pool was developed across tracheotomy, ventilator, and BiPAP/ CPAP specialties
- Content validation by CMH expert panel and clinical staff review
- Caregiving forms were assembled for separate medical technologies
- Field-test of pilot forms
- Data transformation to linear scales with a mathematical measurement
- Instrument refinement and item bank consolidation
- Co-calibration of separate caregiving survey forms

CAMT mothers assume a variety of tasks and duties caring for their children at home. Over three hundred items were initially developed to address this content structure. Professional literature was consulted to identify range of duties and responsibilities. Then content and clinical procedures were conducted to verify correspondence between items and content categories, which are described below.

Content validation was conducted by three independent expert judges. The protocol required classifying each item for a) priority, b) Functional Caregiving, and c) CAMT group. Based on this review, any item that two judges did not agree upon was eliminated from the CAMT item bank. Three judges agreed on 78 percent of CAMT item content classification. This initial reduction of the CAMT item bank established content validity.

After content validation, all items were submitted to a clinical staff review. Four clinical staff persons in Pulmonary/ Allergy/Transitional Care Unit at Children's Memorial Hospital rated each item for priority to CAMT caregiving. Only items rated high or medium priority by majority of clinical staff were retained for form development.

Three forms are the product of this effort and presented below:

- *Tracheostomy* (40 items)
- *Tracheostomy and Ventilator* (36 items)
- *BiPAP/CAP by mask* (31 items)

The following rating scale was implemented to collect responses from mothers:

1 (*None*), 2 (*A little*), 3 (*Somewhat*), 4 (*A lot*), 5 (*Completely*).

In addition, each questionnaire collected information about demographic background, health status, and mental depression. Several anchor items (N = 16) were also included from the Functional Caregiving literature to assess comparability of CAMT and FC caregiving measurement structures.

2.3 Procedure

A list was prepared of children recently receiving medical treatment at CMH and now assisted with medical technologies (ventilator, tracheostomy or BiPAP/CPAP) at home. Then forms were prepared and mailed to all mothers.

2.4 Analysis

Initial statistical analyses were conducted of forms returned by mothers caring for children with tracheostomy, the largest group responding to the survey. Then a co-calibration procedure was conducted for *Tracheostomy*, *Tracheostomy and Ventilator*, and FC Core items. First, raw score means, *SD*s, item difficulties, inter-item correlations, and internal-consistency reliability were computed. Then Principal Components Analysis (PCA) was conducted of raw scores before transformation with a measurement model and co-calibration. Unplanned or large, multiple factors revealed by PCA may introduce statistical dependencies among items, which disturb computation of measurement model fit statistics. Consequently, PCA was conducted to monitor underlying data structure both before and after parameter estimation, which provides a foundation for assessing dimensionality threats. The measurement dimension and raw score PCA, logically, should show substantial consistency.

A Rasch model for rating scales was implemented to transform caregiving raw ratings to linear measures on an objective, equal interval scale (Wright and Masters, 1982; see also Bezruczko, 2005). Raw ratings of mothers and ordinal item scores were transformed to linear units (log-odds) with WINSTEPS software for Rasch measurement (Linacre 2006). The measurement model then computes differences between items and mothers that are also guided by the one-parameter logistic function. The mathematical model for these raw data transformations and representation on a common scale is:

$$\Pi_{nix} = \frac{\exp \sum_{j=0}^{x} [\beta_n - (\delta_i + \tau_j)]}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} [\beta_n - (\delta_i + \tau_j)]} \qquad (1)$$

Where β = mother's confidence, δ = task difficulty, and τ = rating scale thresholds. Π_{nix} is the probability that any item δ_i will be rated X by mother β_n where X takes a value from a fixed range (j = 1, 2, 3, 4, 5), m = number of steps for an item, and k = ith step.

After examining rating scale use, stability of item parameter estimation was compared between older and younger mothers, as well as several other characteristics: income, race, education, employment, and years assisted by medical technology.

The FC theoretical model consists of three levels of caregiving: Advocacy, Personal Caregiving, and Community Relations and relevant items. Construct validity was investigated by statistical decomposition of calibrated item difficulties into caregiving categories that were hypothesized to underlay the FC construct. While Advocacy, Personal caregiving, and Community relations are fundamental to caregiving, item decomposition provides evidence for their hierarchical statistical relations. Several methods are appropriate for this analysis (Embretson and Reise, 2001), and we adapted a multiple regression method described by Green and Smith (1987). Bezruczko and Chen (2007) effectively implemented multiple regression-based item decomposition to investigate construct validity of a FC scale for intellectually disabled adult-children. In this research, item difficulties were decomposed with the following hierarchal content model

$$Y = \text{Advocacy} + \text{Personal Caregiving} + \text{Community Relations} \quad (2)$$

where item caregiving codes were regressed in block fashion, first Advocacy, then Personal Caregiving, and, finally, Community Relations on calibrated item difficulties, Y. The hierarchical block order represents theoretical assumptions about relative importance of item content level for item difficulty.

Another concern was dimensionality of the obtained empirical construct, which was based on an analysis of model residuals. Residuals are differences between observed ratings and model expectations, and depending on category number, an expected value in the rating scale model is given by

$$E_{ni} = \sum_{k=0}^{m} k\Pi_{nik} \quad (3)$$

where Π_{nik} is mother n's probability of responding in category k to item i. For each category the model predicts a response, which is summed across categories. When the expected value E_{ni} is subtracted from the observed response X_{ni}, a score residual Y_{ni} occurs.

$$Y_{ni} = X_{ni} - E_{ni} \quad (4)$$

Then variance of X_{ni} is given by

$$W_{ni} = \sum_{K=0}^{m} (k - E_{ni})^2 \Pi_{nik} \quad (5)$$

and a standardized residual by

$$Z_{ni} = Y_{ni} (W_{ni})^{1/2} \quad (6)$$

In this research, item and person fit to the model were assessed with standardized INFIT and OUTFIT statistics, while unstandardized residuals were examined with PCA. In order to assess dimensionality threats, item residuals were examined with principle components factor analysis.

Congruence or fit of empirical data to this model formulation is established by comparing predicted values and observed values (O-P), which follow a Chi-square distribution and are evaluated with standardized fit statistics for both caregivers and items. In the simple dichotomous case where O_{ni} is the interaction of caregiver n with qualitative indicator i and can take the value 0 or 1, the probability p_{ni} for any observation O_{ni} (0,1) is

$$p_{ni} = \exp(b_n - d_i)/((1 + \exp(b_n - d_i)) \quad (7)$$

where, b_n = estimated position of caregiver n on the FC construct dimension and, d_i = estimated calibration of qualitative indicator i on the FC dimension. Then p_{ni} provides an expectation (E) for any particular $O_{ni.}$ The variance of O_{ni} can be estimated with $p_{ni}((1-p_{ni})$ which can be used to standardize any residual ($O_{ni} - E_{ni}$) as

$$z = (O_{ni} - E_{ni})/((E_{ni}(1 - E_{ni}))^{1/2} \quad (8)$$

This estimated residual z is expected to distribute normally with mean = 0 and standard deviation = 1 and was submitted to PCA factor analysis. Satisfactory empirical correspondence between indicators and mathematical measurement model verifies the stochastic foundations for representing qualitative indicators on an equal-interval scale.

Finally, *Tracheostomy*, *Tracheostomy and Ventilator*, and FC core items were co-calibrated to investigate effectiveness of a single, overall construct for CAMT caregiving measurement. Measurement properties including dimensionality analyses were conducted of the co-calibrated structure following procedures described in Bezruczko and Chen (2007).

3. RESULTS

Distribution of returned questionnaires was 29 *Tracheostomy*, 19 *Tracheostomy and Ventilator*, and 5 *BiPAP* from 138 forms mailed to mothers, which is a modest response rate but comparable to CAMT published studies. Median age of mothers was 42 years (youngest mother was 22 and oldest was 63 years). Mothers reported the following family ethnic background, white 55.8 percent, Hispanic 9.6 percent, African American 19.2 percent, and 15.4 percent reported mixed ethnic background. The sample was predominantly English speaking, 86.5 percent reported excellent English. Mothers reported the following family income: 16.3 percent less than $20,000, 30.6 percent between $20,000 and $59,000, 32.7 percent between $60,000 and $100,000, and 20.4 percent reported more than $100,000 family income. Martial status: 54.7 percent married, 20.8 percent divorced or separated, remaining mothers reported other (20.8 percent). Many mothers reported fulltime employment (42.3 percent), 21.2

percent part time employment, 34.6 percent unemployed, and 1.9 percent were retired.

The children receiving homecare were almost evenly divided between boys and girls. Median age was 7 years; maximum age was 23 years. Average time assisted by medical technologies was 7 years; maximum time assisted was 20 years. Most children were reported to have 3 or more medical diagnoses (80.9%). Majority of children were receiving private duty nurse (60.4%), and 49.1% reported state sponsored programs, and 5.7% health insurance paid for the nurse.

3.1 Raw Data Summary

Tracheostomy ratings were summed for each mother (N = 29), overall raw score mean = 180.72, *SD* = 15.72, and median was 184. (In order to conduct raw score analyses, item means were substituted for missing values.) Item item means ranged between 3.07 and 4.92 raw score units, and overall item mean = 4.52, *SD* = .34. *Tracheostomy* results show substantial skewing of raw totals toward higher confidence of mothers caring for CAMT. However, only two mothers were above the questionnaire ceiling and none below the floor. Items showed positive inter-item correlations and high score reliability (>.90). The item easiest on the *Tracheostomy* form for mothers to endorse was Item 13 "Prepare my child for returning to school", which was endorsed by most mothers. The hardest item to endorse was Item 31 "Trust others involved in my child's care", which was endorsed by few mothers. *Tracheostomy* inter-item correlations for 40 items was between -.68 and .97. Mean item-total correlation was .38 (*SD* = .44). Alpha reliability for 40 items was .91. PCA of 40 *Tracheostomy* items revealed several components above the criterion eigenvalue, and components 1 (10.73 eigenvalue units, 26.82%), 2 (6.18 eigenvalue units, 15.5%), and 3 (3.57 eigenvalue units, 8.93%) accounted for 51.2 percent of raw score variance. An examination of the component matrix revealed Factor 1 consists of a broad range of caregiving tasks and responsibilities with positive loadings. Factor 2 is a bipolar factor with positive items consisting of child-centered caregiving and items with negative loadings associated mainly with machine operation and medical technology. Factor 3 was not interpretable.

3.2 Raw Data Transformation to Linear Units

The following sections present results from transforming *Tracheostomy* responses to linear measures conducted with WINSTEPS software (Linacre, 2006), which implemented a Rasch model for rating scales (Wright and Masters, 1982). These results show mothers were 1.57 logits (*SD* = 1.91) above the scale mean (including mothers with extreme ratings), which indicate this sample of mothers found most items easy to endorse. However, the large *SD* (1.91) indicates much variability among mothers. Average standard error of measurement was .56, which indicates high item precision relative to sample variability. Consequently, *Tracheostomy* items successfully separate mothers into high and low levels of caregiving (Separation 2.66) and reliability was .86 to .88. The small sample size (N = 29), which tends to diminish item separation and reliability, which was lower, .74. Targeting, item density, scale continuity, precision, reproducibility, rating scale function, and Differential Item Function for the *Tracheostomy* form are described in other reports (see Bezruczko, Chen, Hill, and Chesniak, 2008).

Results showed a need for rating scale optimization. Mothers responding to *Tracheostomy* caregiver questionnaire items were presented a five category rating scale: 1 (None), 2 (A little), 3 (Somewhat), 4 (A lot), and 5 (Completely). Most mothers, however, only responded in categories 4 and 5 (88 percent), which leads to inefficient category thresholds estimation on the measurement scale. Consequently, in these data, categories were recoded 4,4,4,4, and 5, which improved scale efficiency and increased Separation reliability of mothers.

Items and mothers showed excellent consistency with expectations of the measurement model. Meansquare fit statistics show both mother responses and item scores were coherent with the measurement dimension derived from the *Tracheosotomy* form. Although relatively few items required deletion, a raw score PCA found three discrete components in the data. Consequently, interpretation of these fit values requires caution.

Although results are very sketchy, the item hierarchy on the obtained item map tends to support Functional Caregiving theory, which predicts caregiving items in the home will be easier to endorse than items involving neighborhood and community resources and collaboration (Chen, Bezruczko, and Ryan-Henry 2006; see also Chen, Ryan-Henry, and Bezruczko, 2004). An unexpected finding is relatively high endorsement of medical technology-related items by many mothers. In fact, the majority of mothers indicated confidence operating medical technology. Figure 1 shows a map of a common measurement dimension for items and mothers.

PCA of the Rasch model residuals showed the measurement dimension accounts for 77.8 percent of the raw rating variance, and several smaller factors account for remainder (11.2 percent). Factor 1 in the residuals only explained 3.8 percent of standardized residuals. However, Factor 1 revealed two groups of items contribute to scale definition of the *Tracheostomy* dimension: a) Technical Machine Knowledge and b) Basic Child Care Items. In Figure 1, Technical Machine Knowledge items define the lower portion of the variable map, which indicates mothers found them relatively "easy" to endorse, while Basic Care Items occur higher indicating higher difficulty. These items suggest CAMT training at Children's Memorial Hospital, which emphasizes competency operating medical equipment, is generally effective as mothers tend to endorse confidence performing them. The Basic Child Care items, which represent confidence about child-centered caregiving are, in fact, more difficult for mothers to endorse. The results, however, show a small group of mothers expressing very low confidence on all items, and they may require intervention. These results from residual analysis should be investigated again as additional data become available. While they represent a minor influence, additional data will clarify the importance of these residual factors for linear measurement. Figure 2 presents results from comparing

item parameter stability across population subgroups, namely, age, race, family income, education, and employment. In every comparison, most items fall within the identity lines but with exceptions. Co-calibration results in Figure 3 verify measures of mothers' confidence were statistically invariant to scale combination.

CONCLUSIONS

Although the sample is small, these results are very promising about operationally defining a construct to measure mothers' confidence providing care to children assisted with medical technologies. The conceptual results generalizing Functional Caregiving from children with intellectual disabilities to CAMT was largely successful. Finally, application of a Rasch measurement model to mothers' self-report ratings to measure confidence on a linear scale, although data was sparse, appears coherent. With additional data and refinement, this instrument should become clinically useful.

ACKNOWLEDGEMENTS

Research was conducted at Children's Memorial Hospital, Chicago, Illinois with support from a Shaw Nursing and Allied Health Research Grant. We are grateful to the Pulmonary/ Allergy/Transitional Care Unit at Children's Memorial Hospital for their cooperation.

REFERENCES

Bezruczko, N. (Ed.). (2005). *Rasch measurement in health sciences.* Maple Grove, MN: JAM Press.

Bezruczko, N. and Chen, S-P. C. (2007). Nonequivalent survey consolidation: An example from Functional Caregiving. *Journal of Applied Measurement, 8,* 336-358.

Bezruczko, N., Chen, S-P, Hill, C., and Chesniak, J. (2008). Measurement of Mothers' Confidence to Care for Children Assisted with Tracheostomy Technology in Family Homes. *Journal of Applied Measurement* (submitted).

Chen, S. C., Bezruczko, N., and Ryan-Henry, S. (2006). Rasch analysis of a new construct: Functional caregiving for adult children with intellectual disabilities. *Journal of Applied Measurement*, 7, 141-159.

Chen, S-P. C., Ryan-Henry, S., and Bezruczko, N. (2004). Functional caregiving: A new construct for mothers' caregiving to adult children with intellectual disabilities. *International Journal of Nursing in Intellectual & Developmental Disabilities. 1, 2,* Article 1. Retrieved March 23, 2005, from http://journal.hsmc.org/ijnidd

Embretson, S. E. and Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum.

Fleming, J., Challela, M., Eland, J., Hornick, R. J. P., & Martinson, I. (1994). Impact on the family of children who are technology dependent and cared for in the home. *Pediatric Nursing, 20,* 379-388.

Green, K. E., and Smith, R. M. (1987). A comparison of two methods of decomposing item difficulties. *Journal of Educational Statistics, 12,* 369-381.

Hill, C. et al. (2008). Development of an Instrument to Measure Mothers' Caregiving for their Children Assisted With Medical Technologies: An Application of Functional Caregiving. Presented at 32nd Annual Midwest Nursing Society Research Conference, Indianapolis.

Linacre, M. J. (2006). WINSTEPS: Rasch measurement software. Chicago: Institute for Objective Measurement.

Thyen, U., Kuhlthau, K., and Perrin, J. M. (1999). Employment, child care, and mental health of mothers caring for children assisted by technology. *Pediatrics, 103,* 1235-1242.

Wright, B. D. and Masters, G. N (1982). *Rating Scale Analysis.* Chicago: MESA Press.

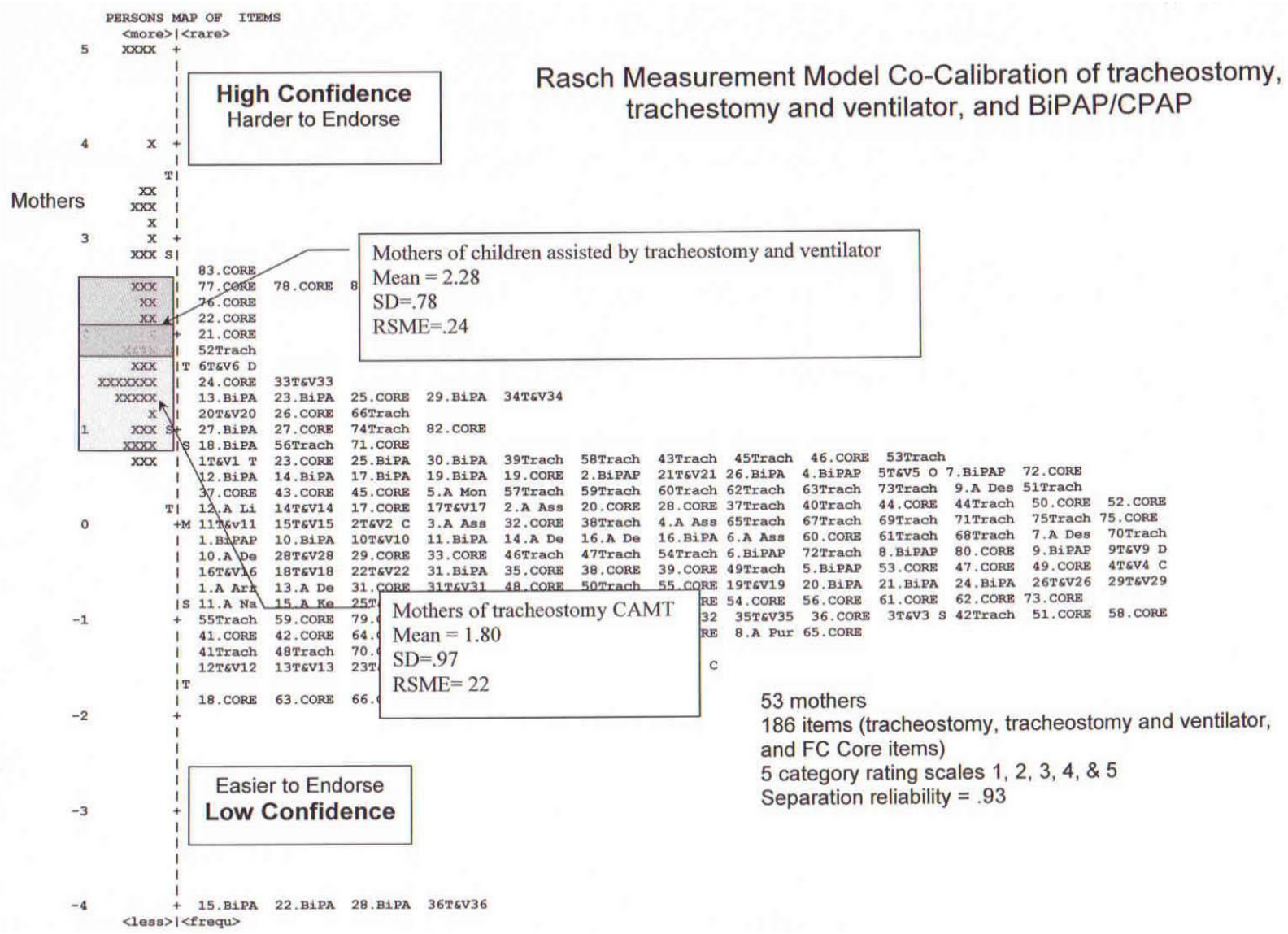

Figure 1. Map of items and mothers after raw scores and ratings were transformed to a linear scale (logits) with Rasch measurement model for rating scales.

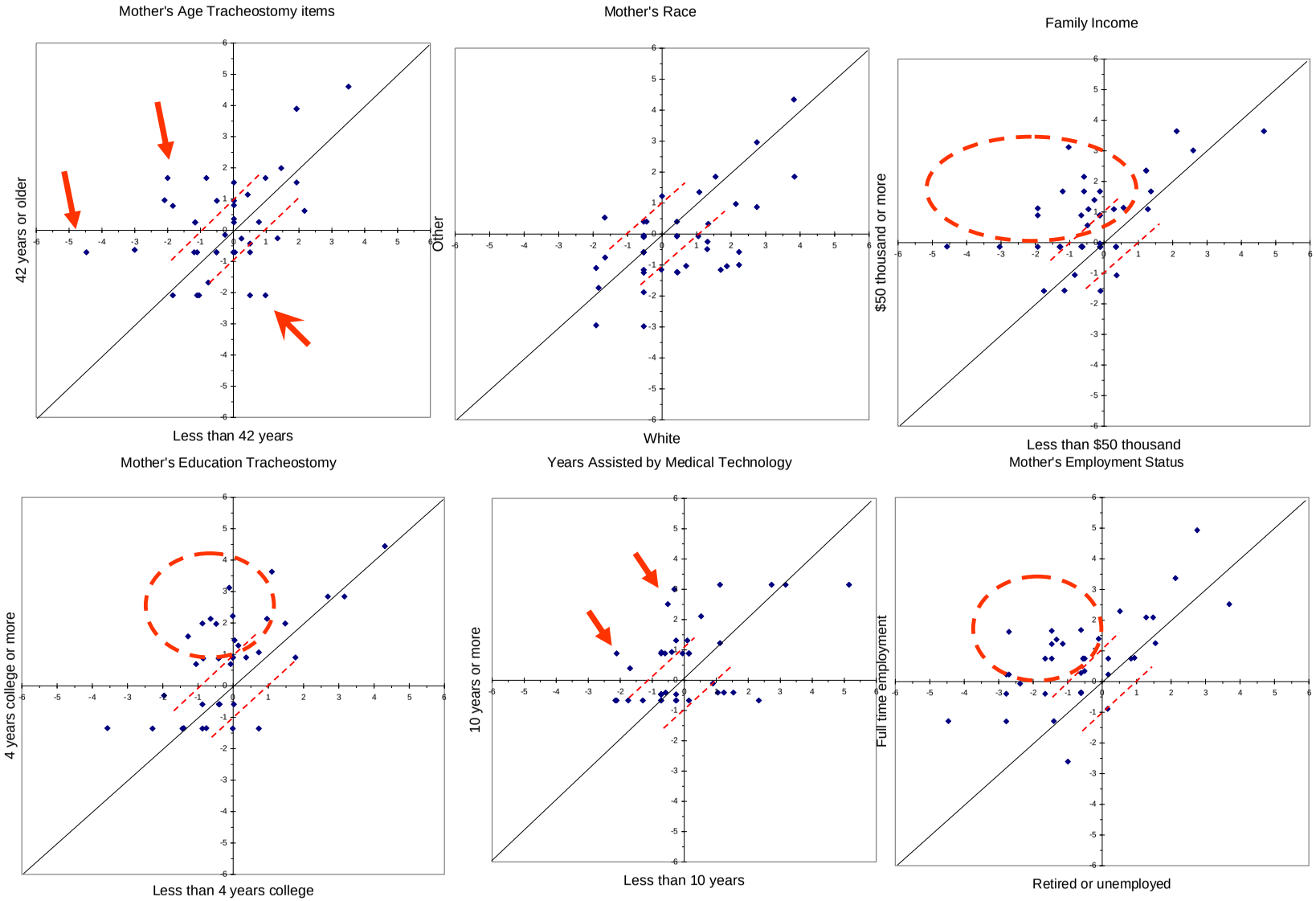

Figure 2. Item parameter stability compared across mothers' age, race, family income, education, years assisted with medical technology, and mothers' employment status.

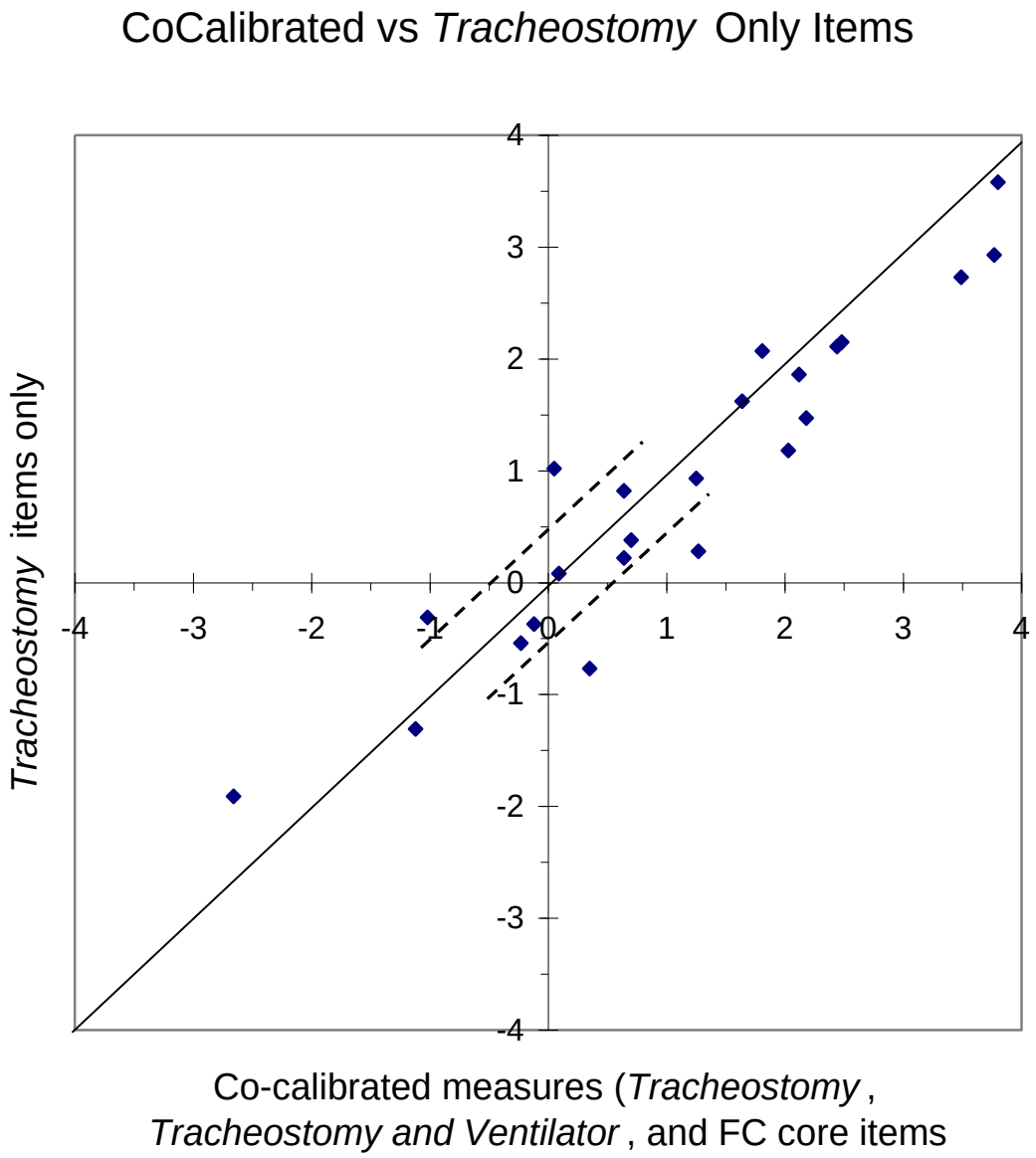

Figure 3. Plot of mothers' caregiving confidence measures after co-calibration of *Tracheostomy*, *Tracheostomy and Ventilator*, and FC core versus *Tracheostomy* Only items.