Gillette Functional Assessment Questionnaire 22-item skill set: factor and Rasch analyses

GEORGE E GORTON III[1] | JEAN L STOUT[2] | ANITA M BAGLEY[3] | KATHERINE BEVANS[4] | TOM F NOVACHECK[2,5] | CAROLE A TUCKER[6]

1 Shriners Hospitals for Children, Springfield, MA, USA. **2** Gillette Children's Specialty Healthcare, St Paul, MN, USA. **3** Shriners Hospitals for Children, Northern California, CA, USA. **4** Children's Hospital of Philadelphia, Philadelphia, PA, USA. **5** Department of Orthopaedics, University of Minnesota, Minneapolis, MN, USA. **6** Temple University, Philadelphia, PA, USA.

Correspondence to George E Gorton III at the Clinical Outcomes Assessment Laboratory, Shriners Hospitals for Children, 516 Carew Street, Springfield, MA 01104, USA. E-mail: ggorton@shrinenet.org

PUBLICATION DATA
Accepted for publication 12th September 2010.
Published online 18th November 2010.

ABBREVIATIONS

DIF	Differential item functioning
FAQ	Functional Assessment Questionnaire
IRT	Item response theory

AIM To determine dimensionality and item-level properties of the Gillette Functional Assessment Questionnaire (FAQ) 22-item skill set using factor and Rasch analyses.
METHOD A retrospective review of parent-reported FAQ 22-item skill set data was conducted of 485 individuals (273 males, 212 females; mean age 9y 10mo, SD 3y 10mo), including 289 with cerebral palsy and 196 with a variety of other neuromusculoskeletal conditions with orthopedic impairments. Factor analyses to validate unidimensionality of the skill set and Rasch analyses to determine relative item difficulty, item and test level information, and content coverage of the item set were performed. Differential item functioning analysis of sub-groups based on sex, diagnosis grouping, and age was conducted. Precision of score estimates for the item set was analyzed.
RESULTS The FAQ 22-item skill set demonstrates unidimensional structure and good item fit statistics. No floor or ceiling effects were noted. Differential item functioning (DIF) based on age was noted for seven items, four items showed diagnosis group-related DIF, and one item sex-related DIF. Precision was adequate in the mid-range range of abilities.
INTERPRETATION Based on this analysis, the FAQ 22-item skill set is a hierarchical set of interval scaled items suitable for measuring locomotor skill ability in children.

The Gillette Functional Assessment Questionnaire (FAQ) is a self or proxy-report measure that includes a ten-level classification of ambulatory function (FAQ Walking Scale), and 22 functional locomotor activities rated on a five-level Likert difficulty scale (FAQ 22-item skill set). The FAQ is intended for use in individuals with all levels of walking ability, and focuses on what an individual can do independently with the use of assistive devices or orthoses as needed to maximize function. The FAQ Walking Scale is a reliable and validated tool that measures functional walking status,[1,2] discriminates among levels of ambulatory ability in cerebral palsy (CP),[3–5] and has been used to monitor changes following treatment.[6–9] The FAQ 22-item skill set was designed to provide further differentiation of higher ambulatory levels and is typically administered with the FAQ Walking Scale. The FAQ 22-item skill set includes common functional mobility activities that further characterize the individual's locomotor abilities.

Recently, physical function outcome assessments have utilized factor analyses along with Rasch or item response theory (IRT) analyses to help develop sets of items, called item banks, whose factor structure and item level statistics are known. These item banks can be used for a number of purposes including computer adaptive testing or tailored short forms. Factor analysis assesses the dimensionality of a set of items relative to an underlying construct such as physical function. Rasch and IRT analyses determine the hierarchical difficulty of individual items and estimates of person abilities on a single, common interval measurement scale which describes the underlying construct. Rasch differs from IRT in the underlying mathematical models, with the most important difference being that only the item thresholds are determined in Rasch-based models. In IRT, two additional parameters can be determined providing a more complex model of the data, but this requires additional data. For healthcare applications, the practical differences between Rasch and IRT models are small. In both analyses, the person abilities and item difficulties are determined relative to the same underlying measurement scale, and the relationships between item and person characteristics can be examined. The overlap between ranges of person abilities and item difficulties (content coverage), and the precision of the estimated person ability scores along the continuum of abilities by the set of items can be determined. Differential item functioning (DIF), how consistently the items estimate person abilities across differing characteristics (e.g. sex, diagnosis, age), is assessed to determine items that may introduce measurement bias in scores for some groups.

 DOI: 10.1111/j.1469-8749.2010.03832.x

To our knowledge, the factor structure and item level analyses of the FAQ 22-item skill set have not been previously reported. The purpose of this report is to determine the dimensionality and item level characteristics of the FAQ 22-item skill set using factor and Rasch analyses to examine DIF, content range and precision of the measure. Analyses of the FAQ using factor and IRT methods would provide an interval level of classification improving its utility as a clinical measure.

What this paper adds

- Dimensionality and item level characteristics of FAQ 22-item skill set were determined using factor analysis and Rasch methods.
- Unidimensionality was confirmed with good item fit statistics, no floor or ceiling effects, age-related DIF in seven items, and no diagnosis or sex-related DIF.
- Precision was adequate centrally but decreased in the extremes.

METHOD

Participants

Retrospective medical record review of FAQ 22-item skill set data was conducted of 485 individuals (mean age 9y 10mo, SD 3y 10mo; median 9y 8mo; range 3.4–19y; 273 male and 212 female). These included 289 with CP in the Gross Motor Function Classification System (GMFCS: 104 level I, 97 level II, 69 level III, 19 level IV), and 196 children with other neuromusculoskeletal diagnoses including developmental (*n*=73), neuromuscular (*n*=39), cerebral vascular accident or traumatic brain injury (*n*=17), joint disorder (*n*=30), genetic (*n*=11), or other (*n*=26) etiologies. Demographic characteristics are summarized in Table I. Walking ability ranged from marginal to functional community ambulators, with 91 (17.5%) routinely using assistive devices for ambulation. All data were obtained by proxy report from a parent or legal guardian as part of the information collected during routine clinical gait analysis in a hospital setting. Waiver of informed consent and Health Insurance Portability and Accountability Act (HIPAA) authorization were obtained for this retrospective analysis from the local Institutional Review Board.

FAQ 22-item skill set

The FAQ 22-item skill set asks the respondent to 'Please rate how easy it is for the patient to do the following activities' followed by brief descriptions of 22 locomotor skills (Table II). The five-level Likert response scale used for the 22 skill items was 'easy', 'a little hard', 'very hard', 'can't do at all', and 'too young for activity'. Those who responded 'too young for activity' were coded as 'can't do at all' for this analysis. Questions to which the respondent did not respond were treated as missing.

Table I: Demographic characteristics

	Overall	Cerebral palsy	Other
n	485	289	196
Sex (%)	M 273 (56) F 212 (44)	M 172 (60) F 117 (40)	M 101 (52) F 95 (48)
Age, mean (SD)	9.9 (3.8)	9.0 (3.8)	11.1 (3.7)
Median (range)	9.6 (3.4–19.0)	8.2 (3.4–18.2)	10.8 (3.8–19.0)
GMFCS (CP only)		I – 104 II – 97 III – 69 IV – 19	
Diagnostic breakdown (%)			Developmental 73 (15.1) Neuromuscular 39 (8.0) CVA/TBI 17 (3.5) Joint disorder 30 (6.2) Genetic 11 (2.3) Other 26 (5.4)

M, male; F, female; GMFCS, Gross Motor Function Classification System; CVA/TBI, cerebral vascular accident/traumatic brain injury.

Statistical analysis

Confirmatory factor analysis (CFA) to assess unidimensionality of the FAQ 22-item skill set was performed with Mplus software[10] using a weighted least squares estimator with standard errors and mean-adjusted χ^2 test statistic that used a full weight matrix with oblique Geomin rotation.[11] Model fit was examined via multiple indices including the Confirmatory Fit Index (CFI), the Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). The CFI and TLI values greater than 0.95, RMSEA less than 0.06, and SRMR less than 0.08 indicate good fit of the model to the data.[12,13]

The Rasch rating scale model was implemented using Winsteps software[14] to simultaneously determine item location (i.e. the difficulty level of each item relative to other items in the scale), person ability (i.e. the ability level of each person relative to other persons in the sample), and item level fit statistics (degree of variation in the responses relative to the predicted responses). We designated items as fitting well if the model information-weighted fit (infit) and outlier-sensitive fit (outfit) statistics were between 0.5 and 1.50.[15] Infit is sensitive to unexpected response patterns when the item is approximately the same difficulty as the person's ability. Outfit is sensitive to unexpected response patterns when an item is very easy or very hard for the person.

A requirement of Rasch modeling is that the response structure be hierarchical, i.e. higher observations correspond to more of the latent variable. We used a point-measure correlation between the set of responses for an individual item and the set of test scores for each person[15] to evaluate whether the coding scheme and person responses are in accord with this requirement. A good correlation (r>0.6) indicates how well an individual item contributes to the overall estimate of person ability.

Estimated discrimination values greater than 1 indicate that an item discriminates between people with high and low ability better than expected for an item at this difficulty.[15] Items with low estimated discrimination were considered items that would not be as useful in discriminating function.

Differential item functioning or item versus person invariance was examined in Winsteps to determine if item responses were independent of interactions between items and the characteristics of the respondents. In other words, DIF implies that

Table II: The Functional Assessment Questionnaire 22-item skill set items are shown in decreasing order of difficulty. Higher `measure' scores represent more difficult tasks. Item model fit statistics are shown

Item	Description	*n*	Measure (logits)	Model SE	Infit msq	Outfit msq	Pt-meas	Estim discr
S21	Ice skate or roller skate	471	2.45	0.08	1.06	0.77	0.75	1.06
S11	Jump rope	477	2.00	0.08	1.14	0.90	0.76	1.00
S19	Ride 2-wheel bike (without training wheels)	478	1.69	0.07	1.71	1.32	0.71	0.85
S13	Hop on right foot (without holding onto equipment or another person)	476	1.29	0.07	1.03	0.85	0.78	1.07
S14	Hop on left foot (without holding onto equipment or another person)	475	1.28	0.07	0.98	0.84	0.78	1.10
S04	Walk up and down stairs without using the railing	476	0.50	0.07	0.69	0.72	0.80	1.27
S07	Runs well including around a corner with good control	478	0.42	0.07	0.94	1.07	0.75	0.89
S22	Ride an escalator, can step on/off without help	470	0.28	0.07	1.06	1.02	0.75	1.13
S12	Jumps off a single step independently	478	−0.02	0.07	1.02	0.87	0.76	1.26
S10	Can get on and off a bus by him/herself	476	−0.21	0.07	1.02	0.95	0.74	1.18
S02	Walk carrying a fragile object or glass of liquid	479	−0.22	0.07	0.74	0.79	0.76	1.14
S06	Runs	478	−0.28	0.07	1.19	1.57	0.66	0.59
S18	Kick a ball with left foot	475	−0.68	0.07	1.10	1.41	0.65	0.55
S20	Ride 3-wheel bike (or 2-wheel bike with training wheels)	470	−0.68	0.07	1.64	1.75	0.60	0.35
S17	Kick a ball with right foot	476	−0.75	0.07	1.08	1.30	0.66	0.65
S16	Step over an object, left foot first	472	−0.78	0.07	0.78	0.86	0.72	1.10
S15	Step over an object, right foot first	472	−0.80	0.07	0.79	0.81	0.73	1.09
S08	Can take steps backwards	477	−0.89	0.07	0.76	0.81	0.72	1.05
S05	Steps up and down curb independently	477	−0.91	0.07	0.82	0.70	0.73	1.34
S09	Can maneuver in tight areas	467	−0.92	0.07	0.77	0.84	0.72	1.01
S01	Walk carrying an object	477	−1.34	0.07	0.75	0.73	0.70	1.19
S03	Walk up and down stairs using the railing	477	−1.43	0.08	0.70	0.74	0.70	1.18

MSQ, mean square; Pt-meas, point-measure; Estim discr, estimated discrimination.

some other factor, related to group membership but different from the latent trait, influences the item responses. DIF analyses were accomplished by calculating item difficulty differences for one group compared to all groups based on respondent characteristics such as sex, age and diagnosis. Diagnosis was examined by comparing those with CP to those with all other diagnoses combined. Approximately 40 participants in each sub-group are generally sufficient to complete DIF analyses, hence further breakdown was not possible because of the small numbers in other diagnostic groups. Age was examined by breaking participants into five groups with equal numbers of participants in each group. Item-characteristic interactions which resulted in a substantive difference of more than 0.5 log-odds unit (logits) and a significant statistical difference ($p<0.05$) were identified as having evidence of DIF.[16]

The precision of the set of items to estimate person ability is quantified by the standard error of the ability score estimate. With 22 items, possible raw scores range from 22 (respondent answers 1 to all questions) to 88 (respondent answers 4 to all questions). For each possible raw score, the standard error was calculated based on the actual scores of the 485 persons for the 22 items. To place the size of standard errors in context, a standard error of 0.32 corresponds approximately to a traditional reliability coefficient of 0.90, and a standard error of 0.22 corresponds to 0.95.[15]

RESULTS

Confirmatory factor analyses of the FAQ 22-item skill set indicates a unidimensional underlying construct. One factor explained more than 95% of the variance, and the ratio of the first two eigenvalues was 13.6:1 indicating dominance of a single factor. In addition, consistently high factor loadings (range of 0.711–0.904) indicate that approximately 84.7% of the common variance is represented by the single latent factor. Three of the four goodness of fit statistics examined demonstrated good model fit (SRMR=0.064, TLI=0.980, CFI=0.982). The RMSEA was 0.160, larger than desired; recent studies have shown this indicator may be sensitive to the distribution of the underlying dataset, with skewness leading to an indication of poor fit.[13] Measures of dimensionality from the Rasch analysis show that 67.1% of the variance was explained by the Rasch dimension, and only 3.8%, with an eigenvalue of 2.6, was explained by the first contrast.

Several pairs of questions had high standardized residual correlations, suggesting some problems with local item dependency. Items 'runs' and 'runs well including around a corner with good control' had a standardized residual correlation of 0.6. Items 'walk up and down stairs using the railing' and 'walk up and down stairs without using the railing' had a standardized residual correlation of 0.32. Additionally, two sets of items that examined right versus left tasks had residual correlations greater than 0.2.

Rasch-based item difficulties ranged from easiest ('walking up and down stairs using the railing') to hardest ('ice skating or roller skating'). Table II contains the items and details of the item level analysis. Item infit and outfit statistics demonstrated adequate fit of all items except one. This item was 'ride 3-wheel bike (or 2-wheel bike with training wheels)'. This item showed high infit (1.64) and outfit (1.75) statistics, and had poor discrimination (0.35).

The Rasch person-item mapping for the FAQ 22-item skill set (Fig. 1) shows minimal ceiling effect and no floor effect; 3.3% of the respondents scored at the highest level, and less than 1% of the respondents scored at the lowest level. The mean of the person ability distribution matches the mean of the item difficulty distribution, with the items distributed across the sample's mean person ability 1 standard deviation of the mean. There are no clear content gaps within the target range of the items.

Differential item functioning was examined for sex, age, and diagnostic category (CP or not). A difference was noted for sex for the item 'jump rope'; it was harder for males to jump rope by 0.55 logits. Also DIF was noted for diagnosis (CP vs not CP) for four items. It was more difficult than predicted by the model for children with CP on items 'runs' (by 0.89 logits), 'runs well including around a corner with control' (by 0.68 logits), 'ride a 3-wheel bike' (by 0.60 logits) and 'ride a 2-wheel bike' (by 0.52 logits). We examined age-related DIF in four groups (3–6y, 6–10y, 10–14y, and 14–19y). Mean ages for each group were 4.9 (n=99, median 5.0), 8.0 (n=156, median 7.9), 11.7 (n=145, median 11.5), and 15.9 (n=85, median 15.7) years. Age-based DIF was noted for the youngest and oldest age groups for six items. Two items, 'jumping rope' and 'ride 2-wheel bike without training wheels' were significantly harder than the model predicted for the younger group compared to overall item difficulty for all age groups. Another two items, 'getting on an escalator' and 'getting on a bus' were significantly harder than the model predicted for the youngest group to perform, and easier than the model predicted for the oldest group to perform. One item ('runs') was significantly easier than the model predicted for the youngest group and harder than the model predicted for the oldest group. Finally, one item ('kick a ball with the right foot') was significantly harder than the model predicted for the oldest group.

Precision of the FAQ 22-item skill set was assessed by calculating the standard error for each ability score level. The standard error of the model for each score along the ability spectrum is represented in Fig. 2. At the high and low end of scoring, where there are few items that measure or provide information at these ability levels, and few people with those levels of ability, the standard errors are, as expected, higher. In the middle of the ability scale, where many questions target that ability level and many respondents have their ability level, the standard errors are low, equivalent to a reliability coefficient of 0.9.

The correlation between responses to an item and the Rasch ability scores for that item were good, ranging from 0.60 to 0.80 (Table II). Estimated discrimination was generally good;

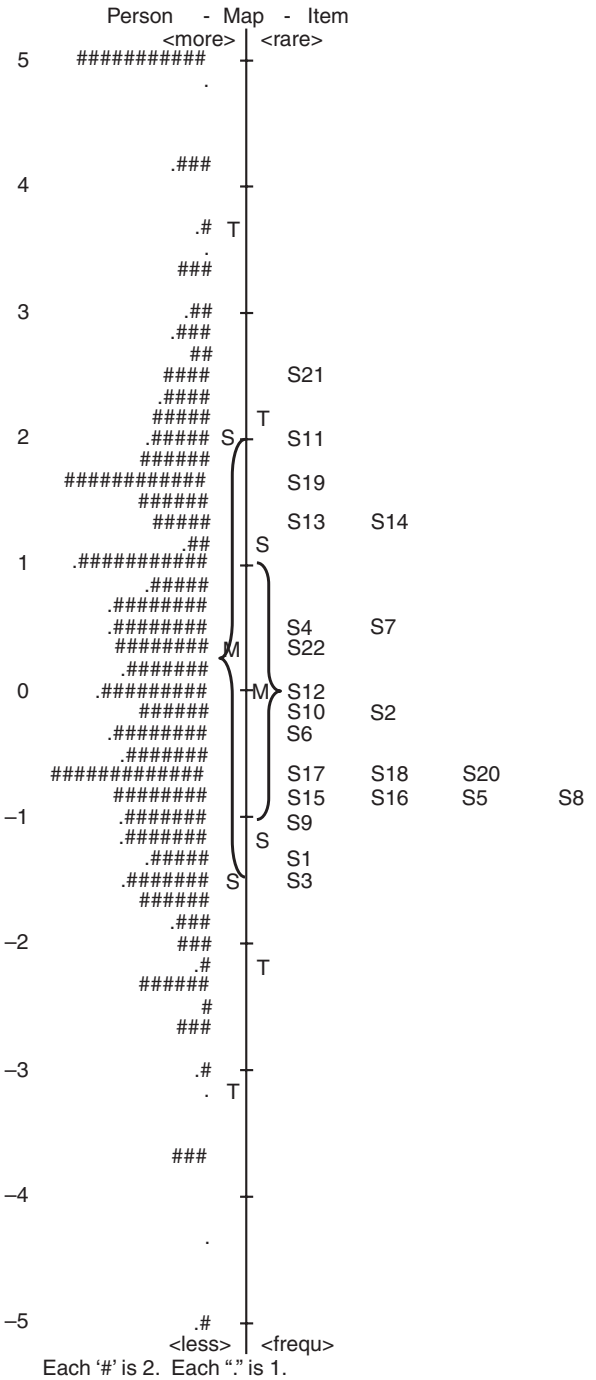

Figure 1: Functional Assessment Questionnaire (FAQ) Rasch Item-Person map. Representation of the frequencies of estimated person ability or `skill' levels for participants compared to the item response levels along the difficulty continuum of lower extremity skills. The vertical scale represents the relative difficulty of physical function with lower numbers representing easier skills, and the higher numbers representing more difficult skills. The left-hand column of # represents the distribution of our sample participants along the skill continuum of lower extremity functioning with each `#' representing 2 participants and each `.' representing 1 participant. The right-hand column represents the distribution of items of the FAQ 22-item skill set. Please refer to Table II for item descriptions. Each side of the vertical scale indicates the mean (M), one standard deviation (S), and two standard deviations (T) for either person-ability or item difficulty (curly brackets). Ideally these two columns should appear similar with the distribution of possible item responses and participants' skill levels spread similarly across the skill scale. Distribution for the 22-item skill set shows that most items surround the mean and one standard deviation.

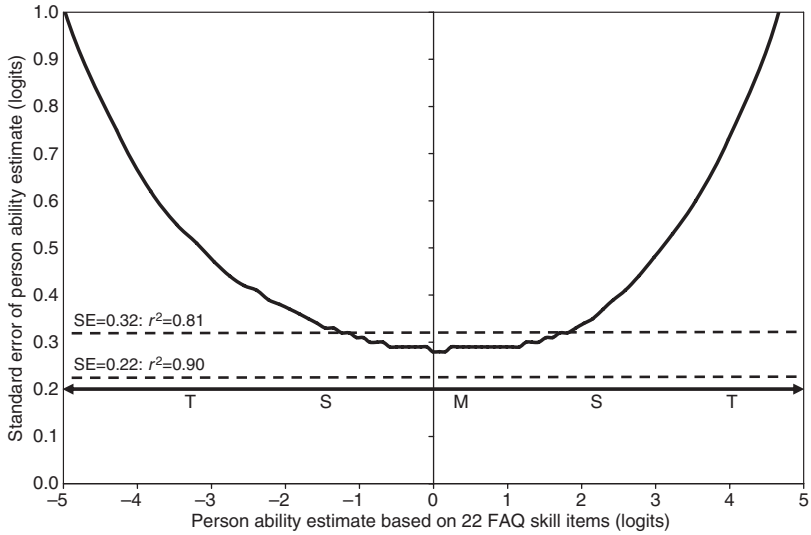

Figure 2: Precision (standard error [SE]) of person ability estimates based on responses to the Functional Assessment Questionnaire 22-item skill set items. Standard errors of less than approximately 0.32 are desirable to match a classical test reliability of r^2=0.81. The distribution of ability for the sample is shown with mean (M) and one and two standard deviations shown (S and T respectively). Precision of the ability estimates is limited further away from the mean because of the limited number of both items and people in those ability ranges.

several items ('rides a 3-wheel bike', 'runs', and 'kicks a ball with right/left foot') had lower discrimination than others of the same difficulty.

DISCUSSION

The FAQ 22-item skill set demonstrates unidimensional structure and acceptable item fit statistics supporting its use as a measure of functional ambulatory skills and as an item bank. While one model fit statistic, the RMSEA, demonstrated inadequate fit, the other model fit statistics from both the CFA and Rasch analyses indicated good model fit, and support the unidimensionality of the item bank.

Other physical functional measures have been subjected to similar analytical approaches. The Gross Motor Activity Estimator (GMAE) is a performance-based 66-item Rasch-based implementation of the Gross Motor Function Measure (GMFM-66).[17–19] The item hierarchies, the ordering of items by difficulty, between the current analysis of the FAQ 22-item skill set and the GMFM-66 are consistent for similar items between the two measures. Hopping on one foot is at the more difficult end of the ability spectrum, and walking is at the easier end. The fewer items in the FAQ 22-item skill set are on the higher ability end of the spectrum of function, and do not cover the basic mobility skills of sitting, standing, kneeling, or crawling as does the GMAE.

The sample size of 485 respondents was sufficiently large to apply a Rasch model. The cross-section of respondents with CP represented 60% of the sample and had a wide range of abilities. The population was slightly biased toward higher functioning individuals as there were fewer low functioning individuals (GMFCS level IV or V) represented in the sample. Further study in a population with lower function (e.g. GMFCS IV) and of older ages (e.g. adults) is warranted to determine the measurement characteristics of this item set.

Difficulty of items in the FAQ 22-item skill set is concentrated around the mean ability level of the population. Within the range of one standard deviation from this mean, the item set, administered as a fixed form, results in acceptable levels of precision. Standard errors of 0.2 to 0.3 are equivalent to a classical test reliability of 0.90 to 0.95. There are insufficient numbers of items at the highest and lowest difficulty levels to ensure high levels of measurement precision for high and low ability individuals. Future work, including assessment of the FAQ 22-item skill set in children with typical development and addition of new items at these levels, should improve the measurement precision.

One item, 'ride 3-wheel bike (or 2-wheel bike with training wheels)' showed poor item fit which may reflect differences in the respondent's interpretation of the type of bike (e.g. adaptive tricycle with full support vs a regular bike with training wheels) and this influenced their response. In addition, a few items of similar tasks, or those that referred to right and left side lower limb performance provided redundant information relative to a specific level of difficulty. Since poorly fitting and redundant items contribute less to the measurement properties, these items could be eliminated with minimal effect on measurement performance. Typically, the redundant item that provides less information is removed.

The DIF analysis showed that several items performed differently for people with different characteristics, such as age, sex, or diagnostic grouping. The items that showed age-related DIF for younger children ('jumping rope', 'getting on an escalator', 'getting on a bus', and 'ride 2-wheel bike without training wheels') were significantly harder than the model pre-

dicted for the younger group (3–6y) and this difference could relate to the limited exposure to these activities or their physical size given their age. Similarly, sex-related DIF for 'jumping rope' may reflect the limited experience males have with this skill in comparison to females. The items 'runs' and 'kick a ball' showed age-related DIF being differentially harder for the younger participants than the model predicted for the oldest group. The reason for this DIF is less clear, perhaps older children attached more performance or skill criteria to 'runs' and 'kicks' than their younger counterparts, interpreting these as more difficult skills in their responses. Finally, several items (the run and bike items) were differentially harder for children with CP. Further exploration of DIF using qualitative research techniques or cognitive interviews would provide a better understanding of the reasons underlying these differences. Using items with significant DIF could result in biased estimates of ability based on the model, and may be considered for removal from the item set when used in large samples or as scales, or may be retained for use in the appropriate subgrouping of individuals.

CONCLUSION

In conclusion, the FAQ 22-item skill set represents a unidimensional, hierarchical set of items with a good range of content appropriate for application to the population of children represented by this sample and may not be generalizable to other populations or to adults. Elimination of items with redundant content, and inclusion of additional high and low ability level items could result in higher precision of estimated ability scores.

REFERENCES

1. Novacheck TF, Stout JL, Tervo R. Reliability and validity of the Gillette Functional Assessment Questionnaire as an outcome measure in children with walking disabilities. *J Pediatr Orthop* 2000; **20:** 75–81.
2. Schwartz MH, Rozumalski A. The Gait Deviation Index: a new comprehensive index of gait pathology. *Gait Posture* 2008; **28:** 351–7.
3. Bagley AM, Gorton G, Oeffinger D, et al. Outcome assessments in children with cerebral palsy, part II: discriminatory ability of outcome tools. *Dev Med Child Neurol* 2007; **49:** 181–6.
4. Oeffinger D, Bagley A, Rogers S, et al. Outcome tools used for ambulatory children with cerebral palsy: responsiveness and minimum clinically important differences. *Dev Med Child Neurol* 2008; **50:** 918–25.
5. Oeffinger D, Gorton G, Bagley A, et al. Outcome assessments in children with cerebral palsy, part I: descriptive characteristics of GMFCS Levels I to III. *Dev Med Child Neurol* 2007; **49:** 172–80.
6. Schwartz MH, Viehweger E, Stout J, Novacheck TF, Gage JR. Comprehensive treatment of ambulatory children with cerebral palsy: an outcome assessment. *J Pediatr Orthop* 2004; **24:** 45–53.
7. Tervo RC, Azuma S, Stout J, Novacheck T. Correlation between physical functioning and gait measures in children with cerebral palsy. *Dev Med Child Neurol* 2002; **44:** 185–90.
8. Trost JP, Schwartz MH, Krach LE, Dunn ME, Novacheck TF. Comprehensive short-term outcome assessment of selective dorsal rhizotomy. *Dev Med Child Neurol* 2008; **50:** 765–71.
9. Stout JL, Gage JR, Schwartz MH, Novacheck TF. Distal femoral extension osteotomy and patellar tendon advancement to treat persistent crouch gait in cerebral palsy. *J Bone Joint Surg Am* 2008; **90:** 2470–84.
10. Muthen BO, Muthen L. Mplus user's guide. Los Angeles, CA: Muthen & Muthen, 1998.
11. Brown TA. Confirmatory factor analysis for applied research. New York NY: The Guilford Press, 2006.
12. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999; **6:** 1–55.
13. Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res* 2009; **18:** 447–60.
14. Linacre J. WINSTEPS Rasch measurement computer program. Chicago: Winsteps.com, 2008.
15. Linacre J, Wright B. A user's guide to WINSTEPS. Chicago: MESA Press, 1999.
16. Zwick R, Thayer DT, Lewis C. An empirical bayes approach to Mantel-Haenszel DIF analysis. *J Educ Measure* 1999; **36:** 1–28.
17. Russell DJ, Avery LM, Walter SD, et al. Development and validation of item sets to improve efficiency of administration of the 66-item Gross Motor Function Measure in children with cerebral palsy. *Dev Med Child Neurol* 2010; **52:** e48–54.
18. Avery LM, Russell DJ, Raina PS, Walter SD, Rosenbaum PL. Rasch analysis of the Gross Motor Function Measure: validating the assumptions of the Rasch model to create an interval-level measure. *Arch Phys Med Rehabil* 2003; **84:** 697–705.
19. Russell DJ, Avery LM, Rosenbaum PL, Raina PS, Walter SD, Palisano RJ. Improved scaling of the gross motor function measure for children with cerebral palsy: evidence of reliability and validity. *Phys Ther* 2000; **80:** 873–85.