Cognitive Functioning Item Bank: Approach to Generating Items for the ASCQ-Me Cognitive Functioning Item Pool[1]

Several physician-researchers on the Adult Sickle Cell Quality of Life Measurement Information System (ASCQ-Me) advisory board asserted that a complete measurement system for sickle cell disease (SCD) should include measuring cognitive functioning. At that time (Spring, 2006), the Patient-Reported Outcomes Measurement Information System (PROMIS®) did not yet have a cognitive functioning item bank, so we proposed to develop one for ASCQ-Me. Our method for generating questions was to be guided by expert opinion, and the results of our literature review rely on information provided by patient interviews and in patient focus groups. Despite using *cognitive* as a keyword, our literature review has not uncovered any extant patient-reported outcomes (PRO) or health-related quality-of-life measures that include questions about cognitive functioning and that were tested in adults with SCD. Moreover, cognitive problems were mentioned only once by one person out of the 120 who participated in interviews and focus groups. Despite concerns that some participants might have cognitive challenges, only once did a moderator feel that a participant had trouble comprehending and expressing himself. Thus, the formative research yielded little material that could be used to generate questions for the ASCQ-Me cognitive functioning item bank.

We based the development of questions for the ASCQ-Me cognitive functioning item bank on the careful and extensive research and analysis conducted by the medical outcomes study (MOS), which is the basis of the MOS patient survey. The MOS identifies memory, attention, executive functioning, and language as aspects of cognitive functioning that should be represented in a self-report questionnaire.[2] Based on this guidance, we generated 28 questions on cognitive functioning for ASCQ-Me. These questions were reviewed by our patient advisory panel, and any potentially problematic questions were also cognitively tested.

Analysis of Field Test Data on the Cognitive Functioning Items

Exhibit 1 demonstrates classical evidence of high internal consistency reliability for the 28 items. Cronbach's coefficient alpha[3] is well above the level that is traditionally considered acceptable for individual-level measurement (> 0.90). Five out of 28 items, however, exhibit item-total correlations that are lower than desirable (< 0.40) (see bold numbers in Exhibit 2). Consequently, these items have been removed from the subsequent analyses.

---

[1] Authors: San Keller, PhD, and Manshu Yang, PhD, September 16, 2015.
[2] Stewart, A. L., Ware, J. E., Sherbourne, C. D., & Wells, K. B. (1992). Psychological distress/well-being and cognitive functioning measures. In A. L. Stewart & J. E. Ware (Eds.). *Measuring functioning and wellbeing: The Medical outcomes study approach* (pp. 102–142). Durham, NC: Duke University Press.
[3] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Exhibit 1: Internal Consistency Reliability for Cognitive Functioning Items

Cronbach's Coefficient Alpha	
Variables	***Alpha***
Raw	0.948650
Standardized	0.958045

Exhibit 2: Item-Total Correlations

Item	**N**	**Item-Total Correlation**
VERBAL_1	556	0.680
VERBAL_2	556	0.701
VERBAL_3	556	0.731
VERBAL_4	556	0.657
VERBAL_5	556	0.643
VERBAL_6	556	**0.275**
VERBAL_7	556	0.787
VERBAL_8	556	**0.376**
VERBAL_9	556	0.673
VERBAL_10	556	0.729
EXECUTIVE_1	556	0.713
EXECUTIVE_2	556	0.793
EXECUTIVE_3	556	0.784
EXECUTIVE_4	556	0.774
EXECUTIVE_5	556	0.805
EXECUTIVE_6	556	0.737
EXECUTIVE_7	556	**0.348**
ATTENTION_1	556	0.746
ATTENTION_2	556	0.738
ATTENTION_3	556	0.795
ATTENTION_4	556	**0.240**
ATTENTION_5	556	0.755
MEMORY_1	556	0.708
MEMORY_2	556	0.747
MEMORY_3	556	0.694
MEMORY_4	556	**0.335**
MEMORY_5	556	0.787
MEMORY_6	556	0.645

We identified the probable number of factors in each item pool by conducting parallel analysis (PA)[4] based on principal axis/common factor analysis on permutations of the raw data (which does not assume normally distributed data). Saturation of the models by a single underlying dimension is described by calculating the ratio of the variance accounted for by the first (superordinate) factor to total observed variance in the data. By comparing eigenvalues from the raw ASCQ-Me data and the 95th percentile of eigenvalues from randomly generated data with the same sample size and numbers of items, we note that the parallel analysis suggests that four factors need to be extracted to account for the covariances between observed cognitive functioning item responses (see Exhibit 3).

Exhibit 3: Results of Parallel Analysis of Cognitive Functioning Item Responses

Root	Raw Data Eigenvalues	Random Data Eigenvalues
1.000000	13.359497	0.493995
2.000000	1.005381	0.417005
3.000000	0.523969	0.360733
4.000000	0.446243	0.312592
5.000000	0.277771	0.273759
6.000000	0.216191	0.235029
7.000000	0.183303	0.200251
8.000000	0.128373	0.166806
9.000000	0.086857	0.135895
10.000000	0.073887	0.106200
11.000000	0.045495	0.075558
12.000000	0.024353	0.050639
13.000000	0.002196	0.021134
14.000000	-0.023667	-0.006181
15.000000	-0.043047	-0.031745
16.000000	-0.058521	-0.059377
17.000000	-0.089754	-0.083588
18.000000	-0.103768	-0.110793
19.000000	-0.110903	-0.136538
20.000000	-0.124502	-0.162823
21.000000	-0.142407	-0.189212
22.000000	-0.143502	-0.219324
23.000000	-0.157236	-0.250886

[4] Horn, J. L. (1965). A rationale and a test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185.

The factors identified by the PA, based on statistical evidence of covariation, could represent artifacts of the question design (e.g., items share the same word pattern). We followed the PA with a bifactor analysis to evaluate the meaningfulness of the four subdomains.[5] The PA output does not describe the relationship of items to factors. Therefore, to specify the bifactor model, we had to conduct an exploratory factor analysis (EFA) to determine which items were related to the four factors. We conducted an EFA using the polychoric correlation matrix of the ASCQ-Me raw data, restricting the number of factors to those revealed by the PA. We used an oblique solution (Promax rotation method) and examined the standardized coefficients of the regression of items onto factors. Three out of 23 items did not have significant loadings on any of the four factors—that is, loadings on all four factors were smaller than 0.40 (data not reported) and were removed from the subsequent analyses. EFA was reconducted based on 20 items. The variance explained by each factor and the resulting standardized factor loadings are shown in Exhibits 4 and 5, respectively.

Exhibit 4: Results from Exploratory Factor Analysis for Cognitive Functioning Domain with 20 Items

Eigenvalues of the Reduced Correlation Matrix: Total = 15.9064501; Average = 0.79532251				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	13.6901300	12.8040200	0.8607	0.8607
2	0.8861101	0.3264597	0.0557	0.9164
3	0.5596504	0.1552968	0.0352	0.9516
4	0.4043536	0.1093745	0.0254	0.9770
5	0.2949791	0.0830574	0.0185	0.9955
6	0.2119217	0.0498942	0.0133	1.0088
7	0.1620275	0.0515611	0.0102	1.0190
8	0.1104664	0.0168604	0.0069	1.0260
9	0.0936060	0.0620490	0.0059	1.0319
10	0.0315570	0.0115402	0.0020	1.0338
11	0.0200168	0.0249880	0.0013	1.0351
12	-0.0049712	0.0109196	-0.0003	1.0348
13	-0.0158907	0.0064015	-0.0010	1.0338
14	-0.0222923	0.0269550	-0.0014	1.0324
15	-0.0492473	0.0170961	-0.0031	1.0293
16	-0.0663433	0.0173509	-0.0042	1.0251
17	-0.0836943	0.0097175	-0.0053	1.0199
18	-0.0934117	0.0094272	-0.0059	1.0140
19	-0.1028390	0.0168398	-0.0065	1.0075
20	-0.1196787		-0.0075	1.0000

[5] Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*, 19–31.

Exhibit 5: Standardized Factor Loadings

Rotated Factor Pattern (Standardized Regression Coefficients)

	Factor 1	Factor 2	Factor 3	Factor 4
…how much trouble …have answering simple questions?	0.29	0.00	0.64 *	0.04
…how much trouble did you have writing simple things?	0.76 *	-0.05	0.33	-0.14
…how much trouble did you have understanding …?	0.59 *	0.08	0.06	0.22
…how much trouble …finding the correct word…?	0.03	0.05	0.09	0.76 *
…how much trouble did you have talking?	0.57 *	0.00	0.21	0.13
…how often did you have a lot of trouble understanding…?	0.66 *	0.05	0.02	0.27
…how often… trouble finding the correct word..?	0.11	0.10	0.03	0.72 *
…how often did you have a lot of trouble writing…?	0.81 *	-0.03	0.05	0.09
…how often…have a lot of trouble making simple decisions?	0.14	0.07	0.69 *	0.10
…how much trouble did you have understanding…?	0.74 *	0.18	0.10	-0.03
…how often did you have trouble figuring out how to…?	0.78 *	0.16	0.04	-0.01
…how often did you have trouble understanding…?	0.69 *	0.24	0.02	0.05
…how much trouble did you have figuring out…?	0.71 *	0.15	0.00	0.08
…how often... trouble paying attention...what you were doing?	0.14	0.34	0.46 *	0.07
…how often did you react slowly to things…?	0.53 *	0.22	0.14	0.04
…how much trouble …remembering things…?	0.07	0.65 *	0.22	0.01
…how often did you forget what you were doing?	0.13	0.65 *	0.14	0.05
…how much trouble …remembering where you put things?	0.00	0.86 *	0.08	-0.03
…how often …a lot of trouble remembering things…?	0.28	0.52 *	-0.06	0.27
…how often… trouble remembering where you put things?	0.06	0.85 *	-0.10	0.06

Values greater than 0.4 are flagged by an "".*

Commonalities among items within a factor:

Factor 1: trouble *and* did you have

Factor 2: remembering

Factor 3: trouble *and* simple

Factor 4: finding the corect words

We conducted bifactor analyses using structural equation modeling (BF-SEM). Two confirmatory factor analyses (CFAs) were conducted using a matrix of polychoric correlations as the input data set. The first CFA modeled each item response as a function of a single general factor and an error term. The second CFA modeled each item response as a function of a single general factor (G), a *nuisance* factor, and an error term. This second CFA is the bifactor model. We compared the standardized regression coefficients associated with the general factor for both models to assess the degree to which the standardized regression coefficients associated with the primary factor differed when the secondary factors were modeled. If they did not differ greatly (in the range of 0.00 to 0.10), then the secondary factors were interpreted as nuisance factors because they described sources of covariation that were negligible.

In our results, the general factor significantly predicts item responses in both models, and the relationship of items to this factor does not vary appreciably whether or not the nuisance factors are included in the model (see Exhibit 6). This stability is indicative of *essential* unidimentionality.[6] We fit to the data a third model that contains a bifactor analysis incorporating the original 23 items. The fit indices are similar to those obtained from the 20-item model.

Exhibit 6: Comparison of Standardized Factor Loadings on General Factor (G) between (1) the Bifactor Model (1 G + 4 Group Factor) and (2) the General Factor Model (1 G Only) With 20 items

						1 G + 4 Group				1 G Only						Different G
v1	=	0.394	*	F3	+	0.756	*	FG	/	0.783	*	FG	+	1	e1	-0.027
v2	=	-0.626	*	F1	+	0.780	*	FG	/	0.814	*	FG	+	1	e2	-0.035
v3	=	-0.001	*	F1	+	0.852	*	FG	/	0.843	*	FG	+	1	e3	0.008
v4	=	0.526	*	F4	+	0.709	*	FG	/	0.714	*	FG	+	1	e4	-0.005
v5	=	-0.167	*	F1	+	0.795	*	FG	/	0.803	*	FG	+	1	e5	-0.009
v7	=	0.016	*	F1	+	0.911	*	FG	/	0.895	*	FG	+	1	e7	0.016
v9	=	0.524	*	F4	+	0.742	*	FG	/	0.746	*	FG	+	1	e9	-0.005
v10	=	-0.286	*	F1	+	0.833	*	FG	/	0.845	*	FG	+	1	e10	-0.013
v11	=	0.632	*	F3	+	0.764	*	FG	/	0.788	*	FG	+	1	e11	-0.024
v12	=	-0.157	*	F1	+	0.909	*	FG	/	0.913	*	FG	+	1	e12	-0.003
v13	=	-0.112	*	F1	+	0.898	*	FG	/	0.899	*	FG	+	1	e13	-0.001
v15	=	-0.019	*	F1	+	0.921	*	FG	/	0.914	*	FG	+	1	e15	0.007
v16	=	-0.010	*	F1	+	0.877	*	FG	/	0.868	*	FG	+	1	e16	0.009
v18	=	0.259	*	F3	+	0.793	*	FG	/	0.813	*	FG	+	1	e18	-0.021
v19	=	-0.027	*	F1	+	0.831	*	FG	/	0.831	*	FG	+	1	e19	0.000
v23	=	0.279	*	F2	+	0.756	*	FG	/	0.778	*	FG	+	1	e23	-0.022
v24	=	0.340	*	F2	+	0.793	*	FG	/	0.814	*	FG	+	1	e24	-0.021
v25	=	0.604	*	F2	+	0.715	*	FG	/	0.752	*	FG	+	1	e25	-0.037

[6] Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*, 19–31.

v27	**=**	**0.188**	*****	**F2**	**+**	**0.868**	*****	**FG**	**/**	**0.868**	*****	**FG**	**+**	**1**	**e27**	**0.000**
v28	=	0.534	*	F2	+	0.712	*	FG	/	0.735	*	FG	+	1	e28	-0.023

23 Items/1G only Model:	RMSEA = 0.15;	CFI = 0.81;	NNFI = 0.79
23 Items/1G+4group Model:	RMSEA = 0.12;	CFI = 0.91;	NNFI = 0.89
20 Items/1G only Model:	RMSEA = 0.16;	CFI = 0.81;	NNFI = 0.79
20 Items/1G + 4group Model:	RMSEA = 0.12;	CFI = 0.91;	NNFI = 0.89

RMSEA stands for the root mean square error of approximation;
CFI stands for the comparative fit index;
NNFI stands for the non-normed fit index.

We replicated the BF-SEM analyses using item response theory graded response models. For the unidimensional model, there is a single latent trait and correlated errors are not allowed. For the bifactor item response theory model (BF-IRT), each item is allowed to have a discrimination parameter on the general factor and one of the group factors. These analyses identified five pairs of items with high local dependence. Therefore, we removed one item from each pair, resulting in a total of 15 items, which we then submitted to analysis. These discrimination parameter estimates for the general factor obtained from both unidimensional and bifactor models are shown in Exhibit 7. The discrimination parameter estimates for most items do not differ between the models. The Pearson correlation and the root mean squared deviation (RMSD) of discrimination parameters between the two models is 0.955 and 0.479, respectively, indicating essential unidimentionality. In addition, the correlation of latent trait scores obtained from the unidimensional and bifactor models is 0.926.

Taken together, these results support the interpretation of the cognitive functioning item pool as unidimensional. Although the PA and EFA suggest some multidimensionality to the data, examination of the semantic commonalities among the items suggests that this covariance is based on similarities in the phrases used across items rather than in the meaning of the items (see last four rows in Exhibit 5). The results of the bifactor CFA and BF-IRT support this interpretation.

Exhibit 7: Item Discrimination Parameter Estimates for the G in Unidimensional Model and Bifactor Model (15 Items)

Observation	Item	Unidimensional	Bifactor	Difference
1	ATTENTION_1	2.361	3.074	-0.713
2	ATTENTION_2	2.714	2.983	-0.269
3	EXECUTIVE_2	4.144	4.468	-0.324
4	EXECUTIVE_3	3.933	4.492	-0.559
5	EXECUTIVE_5	4.460	4.791	-0.331
6	EXECUTIVE_6	3.416	3.671	-0.255
7	MEMORY_2	2.409	3.113	-0.704
8	MEMORY_5	3.233	4.248	-1.015
9	MEMORY_6	2.014	2.598	-0.584
10	VERBAL_1	2.127	2.497	-0.370

Observation	Item	Unidimensional	Bifactor	Difference
11	VERBAL_10	2.909	3.132	-0.223
12	VERBAL_3	2.851	3.078	-0.227
13	VERBAL_5	2.598	2.755	-0.157
14	VERBAL_7	3.875	4.143	-0.268
15	VERBAL_9	2.013	2.261	-0.248

Correlation = 0.955; RMSD = 0.479

Validity and Reliability Analysis of Cognitive Functioning Item Bank

After fitting IRT models (e.g., graded response model) to each of the six ASCQ-Me item banks, we evaluated psychometric properties of each item bank, including efficiency, reliability, and validity. Exhibits 8–13 show the information curves and standard error curves, which are mirror images of each other for the corresponding item bank and across different levels of the health scores.[7] The data provided by the items in the bank are most precise at the highest level of the information curve and the lowest level of the error curve.

Compared to the other item banks, the social functioning and sleep items provide precise measurement for the widest range of health, covering most of the top and bottom of the continuum. The emotional distress, pain, and stiffness domains provide precise measurement for those who are more burdened with these symptoms than the average of patients who responded. As is typical of symptom measures, these item banks provide limited information for adults who are very healthy: As would be expected, among people reporting little or none of the symptom, it would be difficult to find evidence of a difference in that symptom. The information curves indicate that the cognitive functioning item bank is not appropriate to use for patients or research participants who have average or above-average health scores compared to the respondents in the field test; that is, very little information about cognitive functioning will be provided for such respondents.

[7] Technically, the health scores in this instance are the *latent trait continuum*, or, the values of *Theta* or θ as specified in the IRT equations, but we use the term *health score* here for simplicity. However, note that when we talk about item banks, we are discussing health scores that are derived from their relationships to a latent trait and expressed in theta values.

Exhibit 8: Cognitive Functioning Measure—Total Information and Standard Error Curves

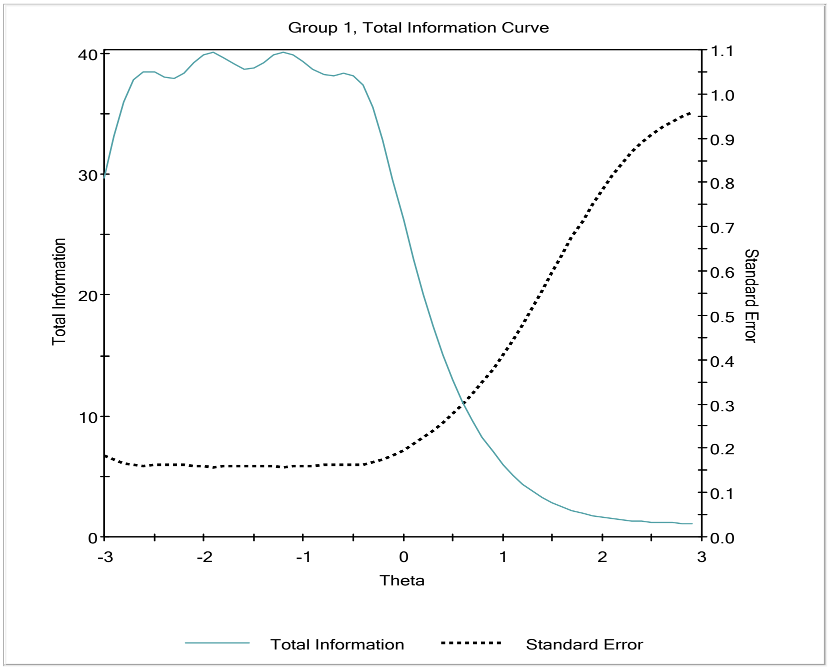


Exhibit 9: Emotional Impact Measure—Total Information and Standard Error Curves

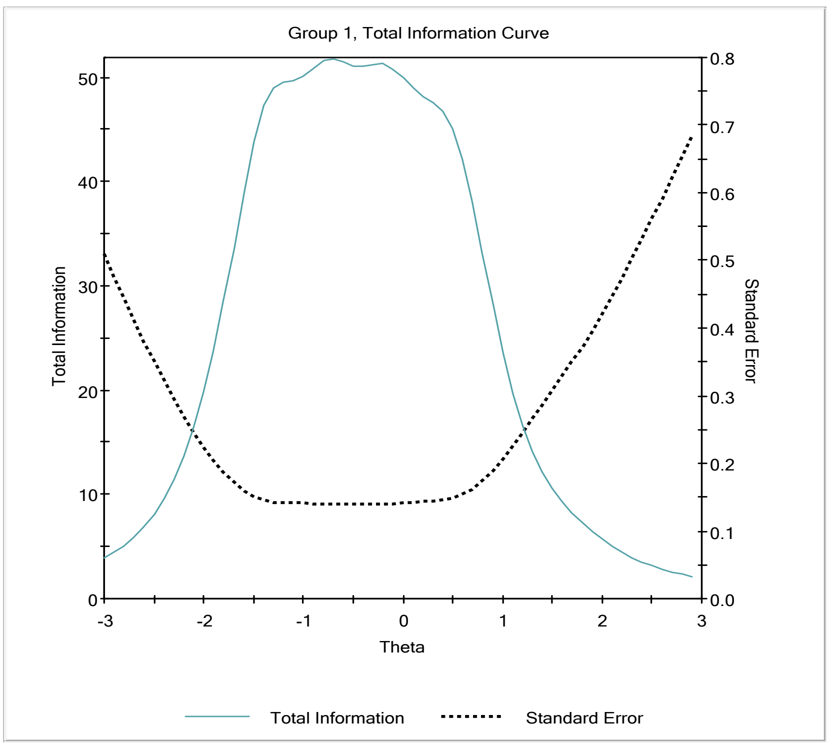


Exhibit 10: Social Functioning Impact Measure—Total Information and Standard Error Curves

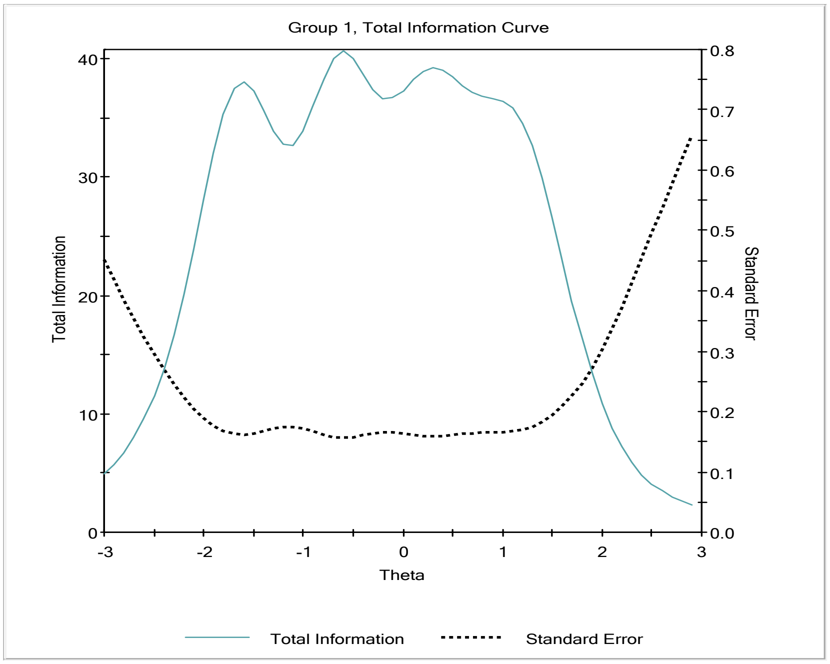


Exhibit 11: Pain Measure—Total Information and Standard Error Curves

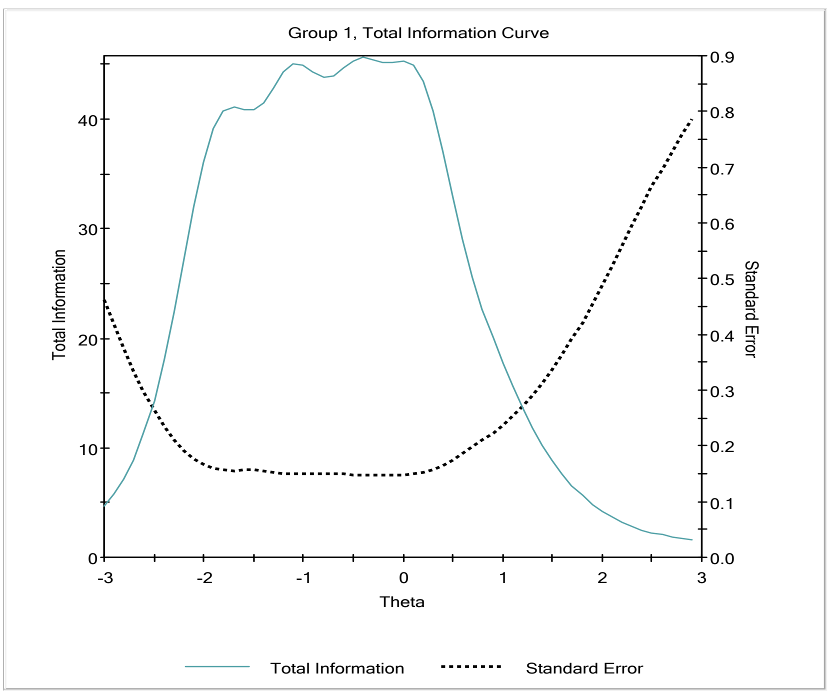


Exhibit 12: Stiffness Measure—Total Information and Standard Error Curves

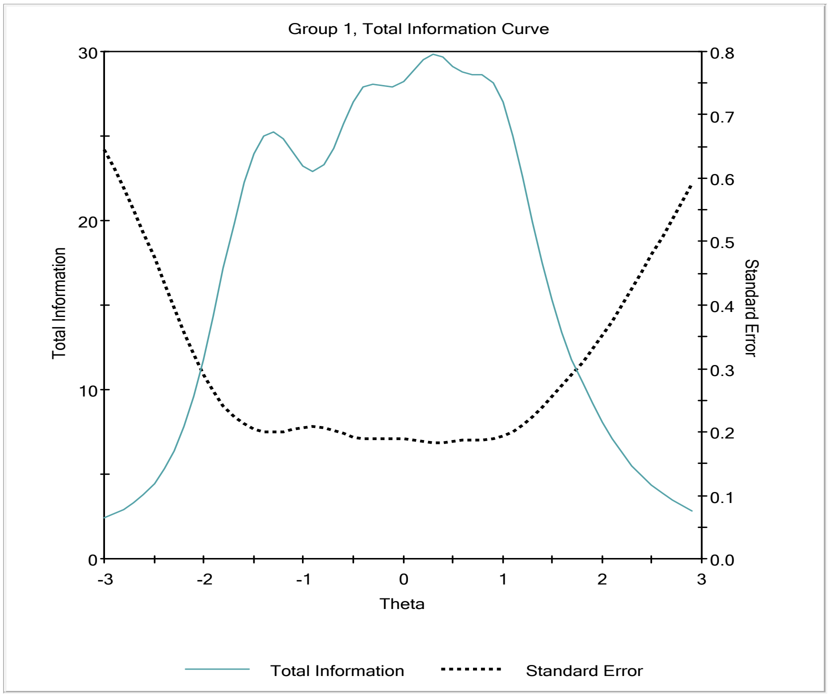
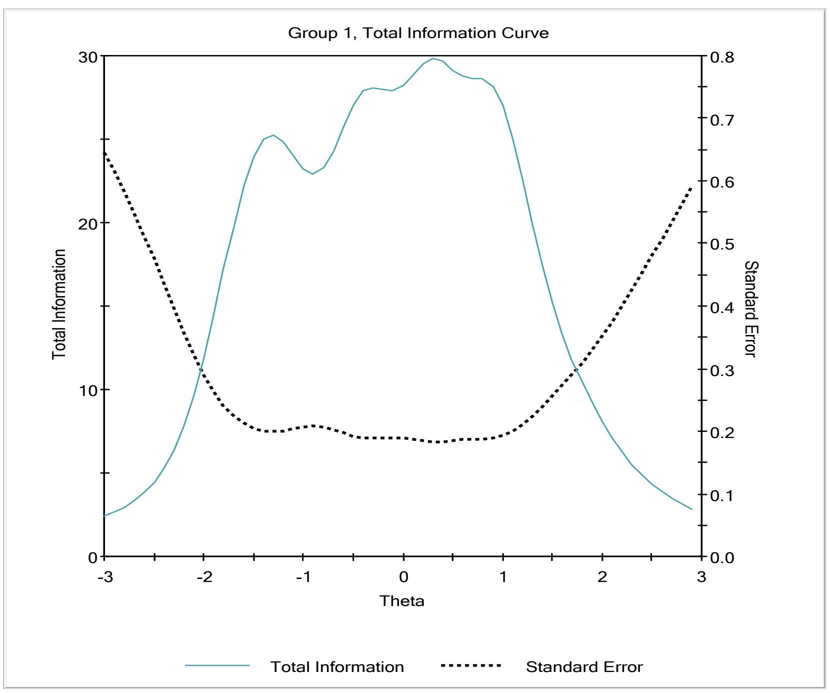


Exhibit 13: Sleep Measure—Total Information and Standard Error Curves



We created graphs matching respondents' scores to item locations (person-item map) for each item bank. With the exception of the cognitive functioning item bank, the scores obtained from each domain appear normally distributed, and location parameters for all the items cover a wide score range. For the cognitive functioning domain (Exhibit 14), however, the person-item map indicates that a large number of potential respondents will be healthier than the highest possible score (i.e., their latent health scores would be > 4.0). Because there is no item located in the upper extreme, very healthy people would not be measured reliably using the cognitive scale. This result is consistent with the total information and error curves for the cognitive functioning measure shown in Exhibit 8.

Exhibit 14: Plot of Relationship Between Respondent and Item Locations

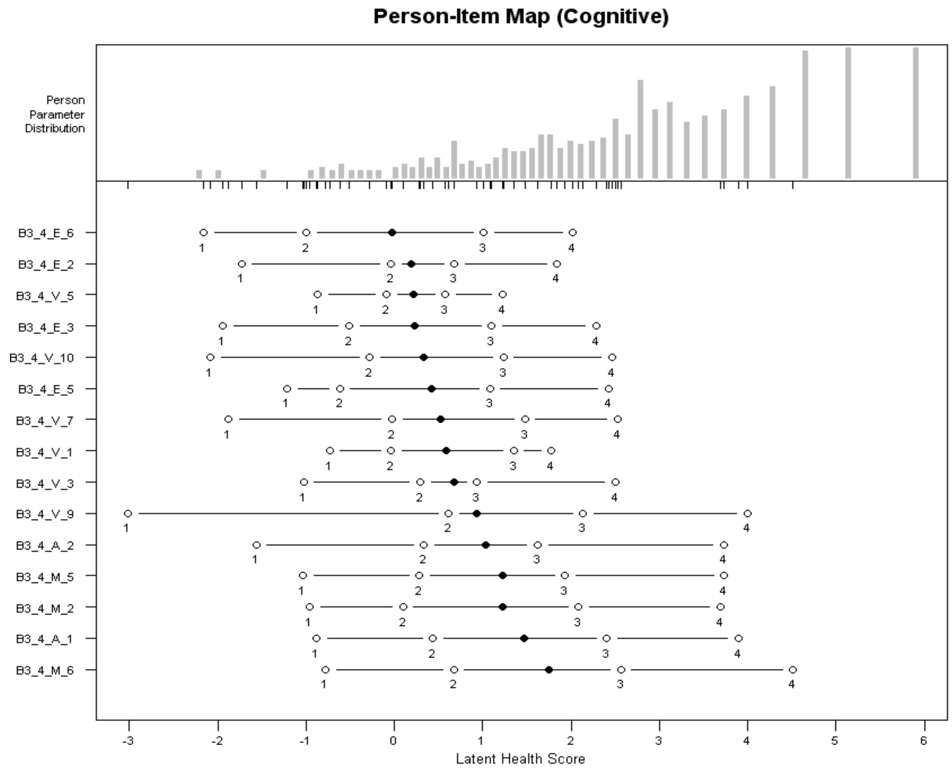


Exhibit 15 shows the correlation coefficients of latent health scores with the SCD severity score (third column) across the six item banks (row headings). The results are sorted in descending order, based on the correlation of the health measure to the indicator of SCD severity. The cognitive functioning item bank has the lowest correlation to SCD severity.

Exhibit 15: Correlation With Severity Levels

Measure Ordered by Relationship With SCD Severity	Number of Items	Correlation With SCD Severity
Stiffness	15	-0.367
Pain	13	-0.344
Social Functioning Impact	17	-0.281
Emotional Impact	20	-0.232
Sleep Problems	12	-0.200
Cognitive Functioning Impact	15	-0.137

Conclusion

Given the weak measurement properties of the cognitive functioning item bank, we did not include it in ASCQ-Me. To assess cognitive functioning based on self-report, we recommend testing the PROMIS applied cognition abilities measures. To assess cognitive functioning based on performance, we recommend testing the National Institutes of Health Toolbox Cognition Battery.