

Evaluation of the Adult Strabismus-20 (AS-20) Questionnaire Using Rasch Analysis

David A. Leske, Sarah R. Hatt, Laura Liebermann, and Jonathan M. Holmes

PURPOSE. To further refine the Adult Strabismus 20 (AS-20) health-related quality of life (HRQOL) questionnaire using Rasch analysis.

METHODS. Rasch analysis was performed independently on the original AS-20 using the following steps: dimensionality, response ordering, local dependence, infit and outfit analyses, differential item functioning, subject targeting, and confirmatory dimensionality.

RESULTS. Two subscales were present in each of the original AS-20 subscales, for a total of 4 subscales, which were labeled “self-perception,” “interaction,” “reading function,” and “general function.” Response ordering was appropriate for 3 of the subscales but required reduction to 4 response options for the fourth subscale. No notable local dependence was found for any subscale. As a result of fit analysis, 2 items were removed, 1 each from 2 subscales. No significant differential item functioning was seen for sex or age. The resulting 5-item self-perception subscale and 4-item reading function subscale are reliable and target the adult strabismus patient cohort appropriately. The resulting 5-item interaction subscale and 4-item general function subscale have less than optimal reliability.

CONCLUSIONS. The AS-20 benefits from reduction to 4 subscales (self-perception, interaction, reading function, and general function) and reducing the response options in the general function subscale from 5 to 4 options. The refined AS-20 may prove to be even more responsive to HRQOL changes in adult strabismus following treatment or changes over time. (*Invest Ophthalmol Vis Sci.* 2012;53:2630-2639) DOI:10.1167/iov.11-8308

The Adult Strabismus-20 (AS-20) questionnaire is a 20-item patient-derived strabismus-specific instrument designed to evaluate health-related quality of life (HRQOL) in adults with strabismus.^{1,2} The AS-20 was created with two distinct subscales (psychosocial and function) and has been shown to be reliable and valid for assessing HRQOL in adult strabismus patients.^{1,3-5} The questionnaire is self-administered and, for each question, patients choose from five Likert-type response

options: “never,” “rarely,” “sometimes,” “often,” and “always.” Each question and subscale of the AS-20 is scored on a 0- to 100-point scale, with 100 points indicating the best quality of life, and is available for download free of charge at www.pedig.net (accessed 2-22-2012).

Traditionally, questionnaires developed without Rasch methods (e.g., classical test theory) give equal weighting to each item, assuming that each item contributes equally (same difficulty) to the overall assessment of the latent trait (e.g., severity of strabismus). In addition, under classical test theory, intervals between response options for each item are uniform, that is, the change between response options at one end of the response option range contributes equally to the latent trait score as a change at the other end of the response option range. Nevertheless, these assumptions of linearity and uniform weighting may not be true, with the difficulty of items and ability of the subjects to endorse an item often differing. Rasch analysis provides a method to appropriately weight responses of each item and rescale an HRQOL instrument to a linear interval-scored instrument by exploring the probability of individual responses in relation to the ability of individual subjects and difficulty of each item on the instrument being used.^{6,7} Both the subject's ability to endorse an item and the item's difficulty are measured on the same scale and expressed as a logit value (logarithm of the odds units). In this way, Rasch analysis has been used to modify and improve existing HRQOL instruments.⁸⁻¹² In the present study, Rasch analysis was used to evaluate and refine the AS-20.

METHODS

Patient Cohort

The AS-20 was completed by 348 adult strabismus patients at the time of their clinic examination in the strabismus practice of one of the authors (JMH). All questionnaires were self-administered without supervision. Median age was 52 years (range: 18 to 88 years). Two hundred and two (58%) were female and 330 (95%) described themselves as “white.” One hundred fifty-three (44%) had undergone previous surgery and 246 (71%) had diplopia at the time of questionnaire administration. Regarding etiology, 154 (44%) were childhood or idiopathic, 114 (33%) were neurogenic, 59 (17%) mechanical, and 21 (6%) sensory. Deviations were primarily exodeviations in 160 subjects (46%, median exodeviation at distance 20 prism diopters [pd], quartiles 14 pd, 35 and 27 pd at near, quartiles 14 pd, 45 pd), esodeviations in 116 (33%, median esodeviation at distance 20 pd, quartiles 10, 30, and 12 pd at near, quartiles 4 pd, 25 pd), vertical deviations in 60 (17%, median vertical deviation at distance 12 pd, quartiles 8, 19, and 12 pd at near, quartiles 5.5 pd, 22.5 pd), torsional deviations in 7 (2%, median absolute torsional magnitude 4°, quartiles 2°, 16°), and postoperative orthotropia in 5 (1%). Median visual acuity was 20/20 (range: 20/15 to 20/63) in the better eye and 20/25 (20/15 to hand motions) in the worse eye. Seventy-six subjects (22%) had ocular comorbidities, such as glaucoma or cataract.

From the Department of Ophthalmology, Mayo Clinic, Rochester, Minnesota.

Supported in part by National Institutes of Health/National Eye Institute Grants EY015799 and EY018810 (JMH), Research to Prevent Blindness, New York, New York (JMH as Olga Keith Weiss Scholar and an unrestricted grant to the Department of Ophthalmology, Mayo Clinic), and Mayo Foundation, Rochester, Minnesota.

Submitted for publication July 28, 2011; revised October 19, 2011 and February 22 and March 8, 2012; accepted March 12, 2012.

Disclosure: **D.A. Leske**, None; **S.R. Hatt**, None; **L. Liebermann**, None; **J.M. Holmes**, None

Corresponding author: Jonathan M. Holmes, Ophthalmology W7, Mayo Clinic, Rochester, MN 55905; holmes.jonathan@mayo.edu.

TABLE 1. The 20 Items of the Original Adult Strabismus-20 (AS-20) Questionnaire

Psychosocial Subscale:
1) I worry about what people will think about my eyes
2) I feel that people are thinking about my eyes even when they don't say anything
3) I feel uncomfortable when people are looking at me because of my eyes
4) I wonder what people are thinking when they are looking at me because of my eyes
5) People don't give me opportunities because of my eyes
6) I am self-conscious about my eyes
7) People avoid looking at me because of my eyes
8) I feel inferior to others because of my eyes
9) People react differently to me because of my eyes
10) I find it hard to initiate contact with people I don't know because of my eyes
Function subscale:
11) I cover or close one eye to see things better
12) I avoid reading because of my eyes
13) I stop doing things because my eyes make it difficult to concentrate
14) I have problems with depth perception
15) My eyes feel strained
16) I have problems reading because of my eye condition
17) I feel stressed because of my eyes
18) I worry about my eyes
19) I can't enjoy my hobbies because of my eyes
20) I need to take frequent breaks when reading because of my eyes

All subjects gave informed consent. Data were collected and analyzed in accordance with the Health Insurance Portability and Accountability Act guidelines and adhered to the tenants of the Declaration of Helsinki.

Rasch Analysis

Since the AS-20 was originally designed to have two distinct dimensions or subscales, (psychosocial and function) these were analyzed in separate Rasch analyses. All analysis was performed using a commercial software program (Winsteps software version 3.72.2, Winsteps Software Technologies, Seattle, WA; available at www.winsteps.com, accessed 2-22-2012). First, the dimensionality of each subscale was analyzed by principal component analysis to determine whether additional dimensions were present, with the goal of avoiding any scoring of unrelated dimensions together. Second, response ordering was checked to determine whether all response options were utilized and interpreted correctly. Local dependence was evaluated to determine whether items functioned independently of one another (i.e., does the response on one item dictate the response on another item). Interitem standardized residual correlations of >0.7 indicated high local dependence (indicating that 50% or more of the variance in the residuals is common between items),¹³ and items showing these levels of dependence were considered for combining or removal. Infit and outfit were assessed and items with mean square infit or outfit values < 0.60 or >1.40 were considered for removal sequentially. In our study, the standardized z -score was not utilized to exclude items because of its dependence on sample size, with elevation of the z -score as sample size increases. The person separation index and the reliability coefficient were evaluated as measures of instrument precision, with a desired person separation index ≥ 2.0 and reliability coefficient ≥ 0.8 .¹⁴ Differential item functioning (DIF) was then assessed for sex and age (\leq median age [52 years] vs. $>$ median age) using the following criteria: a DIF contrast < 0.5 logits defined as

small or absent, DIF 0.50 to 1.0 as minimal (inconsequential), and DIF > 1.0 as notable.¹⁵ Targeting was assessed to determine whether there was any mismatch between the severity of the disease (ability) and level of item discrimination (difficulty). Optimal targeting is characterized by a difference of <1.0 logits between mean person and item measures.¹¹ Test information curves were plotted following item removal as an indication of subscale reliability across the spectrum of strabismus severity. Finally, unidimensionality in the revised subscales was reconfirmed.

RESULTS

Dimensionality

Analyzing the 10 items of the AS-20 psychosocial subscale (Table 1) for dimensionality, 71.3% of the raw variance was explained by the measures (Table 2). Looking within the unexplained variance, 7.8% of the overall variance was explained by the first contrast, with an eigenvalue of 2.7, suggesting a second dimension, with items 1, 2, 3, 4, and 6 (relating to self-perception) loading on the first contrast. The remaining items (5, 7, 8, 9, and 10), were related to interactions with others. Separate Rasch analyses were then performed on the two identified psychosocial subscales, which we labeled self-perception and interaction.

Analyzing the 10 items of the AS-20 function subscale (Table 1) for dimensionality, 57.4% of the raw variance was explained by the measures (Table 2). Looking within the unexplained variance, 10.2% of the overall variance was explained by the first contrast, with an eigenvalue of 2.4, suggesting a second dimension, with items 12, 13, 16, 19, and 20 (relating to reading function) loading on the first contrast. The remaining items (11, 14, 15, 17, and 18) were related to general function. Separate Rasch analyses were then performed on the two identified function subscales, which we labeled reading function and general function.

Response Ordering

Responses to each item within the self-perception subscale were properly oriented (Fig. 1A), indicating proper use and interpretation of each response category. All five response options were therefore retained within the self-perception subscale.

Responses to each item within the interaction subscale were also properly oriented (Fig. 1B), and therefore all five response options were retained for the interaction subscale as well.

Responses to each item within the reading function subscale were properly oriented (Fig. 1C), and therefore all five response categories were retained within the reading function subscale.

When analyzing responses to each item within the general function subscale, it was evident that the "rarely" response option was underutilized for each of the five items (Fig. 1D). Therefore, "rarely" was combined with "never," creating four possible response options. Reanalyzing function subscale responses using four categories, responses for each of the five items were properly oriented and equally distributed (Fig. 1E). All subsequent analyses were performed using four response options within the general function subscale.

Local Dependence

When analyzing local dependence, all in-trait correlations of standardized residuals were between 0.08 and -0.46 on the self-perception subscale and 0.01 and -0.39 on the interaction subscale (Table 3). In-trait correlations of standardized

TABLE 2. Dimensionality Analysis of the Original AS-20 Psychosocial Subscale and Function Subscale

	Psychosocial Subscale		Function Subscale	
	Eigen	%	Eigen	%
Total raw variance	34.8	100.0	23.5	100.0
Explained by measures	24.8	71.3	13.5	57.4
Explained by persons	14.6	41.9	7.1	30.4
Explained by items	10.2	29.4	6.3	26.9
Total unexplained	10.0	28.7	10.0	42.6
1st contrast	2.7	7.8	2.4	10.2
2nd contrast	1.4	4.1	1.5	6.3
3rd contrast	1.2	3.6	1.3	5.5
4th contrast	1.0	2.8	1.1	4.6
5th contrast	0.9	2.7	1.0	4.3
	Item	1st Contrast	Item	1st Contrast
	Number	Loading	Number	Loading
	1	0.64	11	-0.14
	2	0.48	12	0.71
	3	0.67	13	0.31
	4	0.60	14	-0.34
	5	-0.51	15	-0.34
	6	0.31	16	0.71
	7	-0.44	17	-0.50
	8	-0.51	18	-0.62
	9	-0.55	19	0.01
	10	-0.39	20	0.63

residuals were between 0.01 and -0.38 on the reading function subscale and 0.12 and -0.45 on the general function subscale (Table 3). These results for the four subscales indicate that there were no high levels of local dependence between items (all intraitem correlations of standardized residuals <0.7).

Analysis of Infit and Outfit

No large infit or outfit errors were found within the self-perception subscale, (within the range of 0.60 to 1.40, Table 4). The person separation index was 3.12, with a reliability coefficient of 0.91.

Analyzing the interaction subscale, there were no large infit or outfit errors and the person separation index was 1.51, with a reliability index of 0.70.

For the reading function subscale, item 19 was removed for infit and outfit errors outside of the range of 0.60 to 1.40. The remaining four items (12, 13, 16, and 20) had appropriate infit and outfit errors (Table 4) and were used in subsequent analyses. After removal of item 19, the person separation index of the reading function subscale was 2.66, with a reliability coefficient of 0.88.

For the general function subscale, item 14 was removed for large infit and outfit errors, leaving items 11, 15, 17, and 18 (Table 4). After removal of item 14, the person separation index of the general function subscale was 1.73, with a reliability coefficient of 0.75.

Differential Item Functioning

For the self-perception subscale, when assessing DIF for sex, item 2 had a contrast difference between males and females of -0.63 logits (minimal DIF), indicating that females reported more of an impact of this item. No items had notable DIF (Table 6). Despite minimal differential functioning of item 2,

this item was retained. When assessing the self-perception subscale for DIF based on age, no DIF was observed (Table 6).

For the interaction subscale, no items had minimal or notable DIF when assessing either sex or age. No items were removed from the interaction subscale.

For the reading function subscale, no minimal or notable DIF was observed for sex (Table 5). Only minimal DIF was noted for age on item 13 (contrast difference of 0.75 logits, Table 6). Despite this minimal DIF, item 13 was not removed.

For the general function scale, only minimal DIF was noted for item 11, with a contrast difference of -0.61. As with the reading function subscale, this item was retained. Minimal DIF was noted again for age on item 11, but the item was not removed.

Targeting

Analysis of targeting for the self-perception subscale indicated that the mean severity discrimination (difficulty) of the items was well matched to the mean severity (ability) of the condition (1.07 ± 3.90 logits for mean person vs. 0.00 ± 0.51 logits for mean item; Fig. 2A). A wide range of severity of the condition was evident, with person logit values ranging from -6.88 to 6.38. The test information curve for the self-perception subscale is shown in Figure 3A.

Targeting for the interaction subscale indicated relatively poor targeting (2.66 ± 2.26 logits for mean person vs. 0.00 ± 0.54 logits for mean item; Fig. 2B). As with the self-perception subscale, a wide range of severity of the condition was evident, with person logit values ranging from -5.11 to 5.18. The test information curve for the interaction subscale is shown in Figure 3B.

Analysis of targeting for the reading function subscale indicated that the mean severity discrimination of the items was well matched to the mean severity of the condition (0.62 ± 3.23 logits for mean person vs. 0.00 ± 0.76 logits for mean item; Fig. 2C). A wide range of severity of the condition was

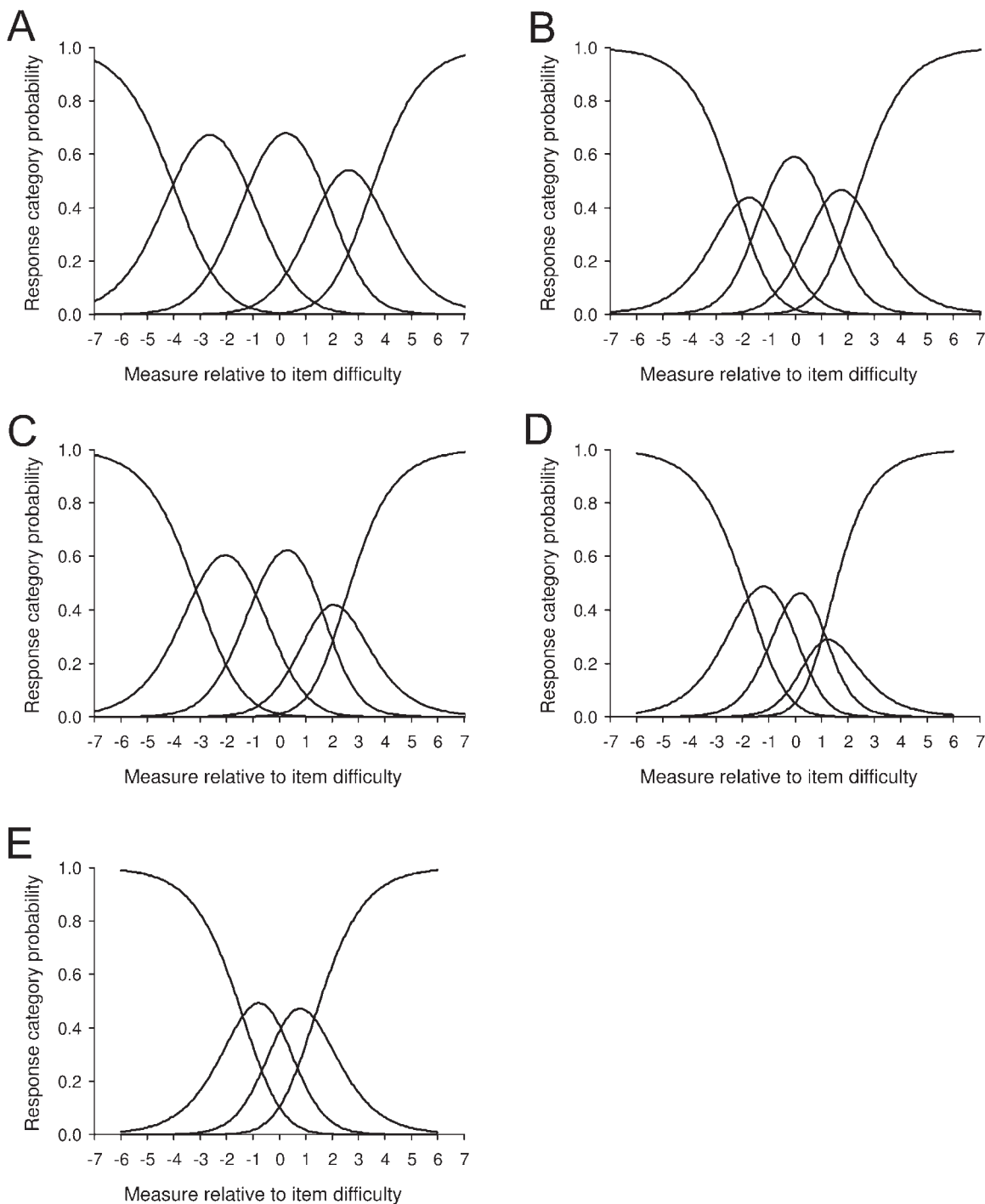


FIGURE 1. Adult Strabismus Questionnaire 20 (AS-20) response ordering for (A) 5-response self-perception subscale (representative item #1), (B) 5-response interaction subscale (representative item #5), (C) 5-response reading function subscale (representative item #12), (D) 5-response general function subscale (representative items #11), and (E) 4-response general function subscale. Response categories were properly oriented and distributed for all items in the self-perception subscale (A), interaction subscale (B), and reading function subscale (C). Using 5 response options for the general function subscale, it was evident that the “rarely” response was underutilized (D). When the “rarely” response option was combined with the “never” response option in the general function subscale, response options were properly oriented and distributed (E). A, always; O, often; S, sometimes; R, rarely; N, never; N/R, never/rarely.

evident, with logit values ranging from -6.75 to 6.01 . The test information curve for the reading function subscale is shown in Figure 3C.

Targeting for the general function subscale indicated appropriate targeting (0.45 ± 2.04 logits for mean person vs.

0.00 ± 0.34 logits for mean item; Fig. 2D). As with the reading function subscale, a wide range of severity of the condition was evident, with person logit values ranging from -4.78 to 4.61 . The test information curve for the general function subscale is shown in Figure 3D.

TABLE 3. Local Dependence Analysis of the AS-20 Self-Perception, Interaction, Reading Function, and General Function Subscales Showing the Highest Correlations between Items

Self-Perception Subscale			Interaction Subscale			Reading Function Subscale			General Function Subscale		
Correlation	Item	Item	Correlation	Item	Item	Correlation	Item	Item	Correlation	Item	Item
0.08	3	4	0.01	7	9	0.01	16	20	0.12	17	18
-0.46	2	6	-0.39	8	9	-0.38	16	19	-0.45	14	17
-0.42	1	4	-0.39	5	10	-0.37	19	20	-0.42	14	18
-0.37	1	3	-0.37	9	10	-0.35	12	19	-0.30	11	17
-0.37	4	6	-0.37	7	8	-0.35	13	16	-0.28	14	15
-0.32	3	6	-0.27	7	10	-0.29	13	20	-0.27	11	18
-0.20	2	3	-0.22	5	8	-0.21	13	19	-0.26	11	14
-0.17	1	2	-0.20	5	7	-0.21	12	20	-0.24	11	15
-0.14	2	4	-0.13	8	10	-0.16	12	13	-0.24	15	18
-0.06	1	6	-0.09	5	9	-0.12	12	16	-0.04	15	17

Confirmation of Dimensionality

Reanalyzing each of the revised subscales for dimensionality revealed that each of the four subscales was unidimensional (Table 7). For the self-perception subscale, 77.2% of the raw variance was explained by the measures, and 8.1% of the unexplained variance was explained by the first contrast, with an eigenvalue of 1.8. For the interaction subscale, 62.0% of the raw variance was explained by the measures, and 12.8% of the unexplained variance was explained by the first contrast, with an eigenvalue of 1.7. The reading function subscale had 73.7% of the raw variance explained by the measures, and 10.4% of the unexplained variance was explained by the first contrast, with an eigenvalue of 1.6. Finally, for the general function subscale, 56.6% of the raw variance was explained by the measures, and 17.7% of the unexplained variance was explained by the first contrast, with an eigenvalue of 1.6.

DISCUSSION

The results of Rasch analysis indicate that the AS-20 questionnaire could benefit from subscale restructuring and reduction of items within the predefined subscales, resulting in a questionnaire with two psychosocial subscales and two function subscales (18 items overall). The new subscales relate to worry about what others think and the perception of self (self-perception, five items), interactions with others (interactions, five items), reading function or near function (reading function, four items), and nonspecific general function (general function, four items). In addition, response options in the general function subscale should be reduced to four options: “Never/Rarely,” “Sometimes,” “Often,” and “Always.”

The person separation index and reliability of the four subscales of the Rasch-revised AS-20 is excellent for the self-perception subscale and the reading function subscale, as indicated by a person separation index ≥ 2.0 and person

TABLE 4. Infit and Outfit Errors of the AS-20 Self-Perception, Interaction, Reading Function, and General Function Subscales, Showing Remaining Items.

Item #	Infit		Outfit	
	Mean Square	Standard Z-Score	Mean Square	Standard Z-Score
Self-perception subscale				
6	1.34	3.7	1.38	3.8
2	0.98	-0.2	1.04	0.5
1	0.93	-0.8	0.91	-1.0
4	0.85	-1.8	0.83	-1.8
3	0.77	-2.9	0.74	-3.1
Interaction subscale				
5	1.09	0.9	1.30	2.2
10	1.16	1.7	1.13	1.4
8	1.03	0.4	1.03	0.3
7	0.83	-1.9	0.87	-1.2
9	0.78	-2.5	0.79	-2.3
Reading function subscale				
13	1.19	2.2	1.29	3.1
20	0.95	-0.6	0.93	-0.9
16	0.94	-0.7	0.92	-1.0
12	0.90	-1.3	0.89	-1.2
General function subscale				
11	1.36	4.3	1.34	4.1
18	0.97	-0.3	0.94	-0.7
15	0.91	-1.2	0.93	-0.9
17	0.74	-3.8	0.70	-4.1

TABLE 5. Differential Item Functioning (DIF) of Sex for the AS-20 Self-Perception, Interaction, Reading Function, and General Function Subscales

Item #	Female		Male		DIF Contrast
	DIF Measure	SE	DIF Measure	SE	
Self-perception subscale					
1	0.09	0.13	0.11	0.17	-0.02
2	-0.52	0.13	0.11	0.17	-0.63*
3	-0.10	0.13	-0.49	0.17	0.38
4	-0.56	0.13	-0.45	0.17	-0.11
6	-1.11	0.13	0.71	0.16	0.40
Interaction subscale					
5	-1.11	0.14	-0.63	0.16	-0.48
7	-0.30	0.13	-0.27	0.16	-0.02
8	0.40	0.12	0.40	0.15	0.00
9	0.30	0.12	0.17	0.15	0.13
10	0.66	0.12	0.41	0.15	0.25
Reading function subscale					
12	-0.96	0.12	-0.78	0.15	-0.18
13	-0.52	0.12	-0.78	0.15	0.26
16	0.77	0.13	0.94	0.15	-0.17
20	0.70	0.13	0.62	0.15	0.08
General function subscale					
11	-0.33	0.12	0.28	0.14	-0.61*
15	0.40	0.11	0.09	0.14	0.31
17	-0.42	0.12	-0.68	0.14	0.26
18	0.36	0.11	0.30	0.14	0.06

* Minimal differential item functioning was found between males and females on items #2 and #11, but these items were retained.

TABLE 6. Differential Item Functioning (DIF) of Age (above and below the Median Age) for the AS-20 Self-Perception, Interaction, Reading Function, and General Function Subscales

Item #	< Median Age		> Median Age		DIF Contrast
	DIF Measure	SE	DIF Measure	SE	
Self-perception subscale					
1	0.11	0.14	0.06	0.15	0.06
2	-0.38	0.14	-0.15	0.15	-0.23
3	-0.12	0.14	-0.40	0.16	0.27
4	-0.50	0.14	-0.55	0.16	0.05
6	0.89	0.14	1.02	0.15	-0.14
Interaction subscale					
5	-1.06	0.13	-0.68	0.17	-0.38
7	-0.23	0.12	-0.42	0.16	0.19
8	0.36	0.12	0.46	0.15	-0.10
9	0.25	0.12	0.25	0.15	0.00
10	0.65	0.12	0.42	0.15	0.23
Reading function subscale					
12	-0.96	0.14	-0.82	0.14	-0.14
13	-0.25	0.13	-1.00	0.14	0.75*
16	0.77	0.13	0.92	0.14	-0.15
20	0.43	0.13	0.92	0.14	-0.49
General function subscale					
11	-0.40	0.13	0.25	0.13	-0.65*
15	0.41	0.12	0.12	0.13	0.29
17	-0.52	0.13	-0.52	0.13	0.00
18	0.49	0.12	0.17	0.13	0.32

* Minimal differential item functioning was found between individuals < median age and individuals > median age for items #11 and #13, but these items were retained.

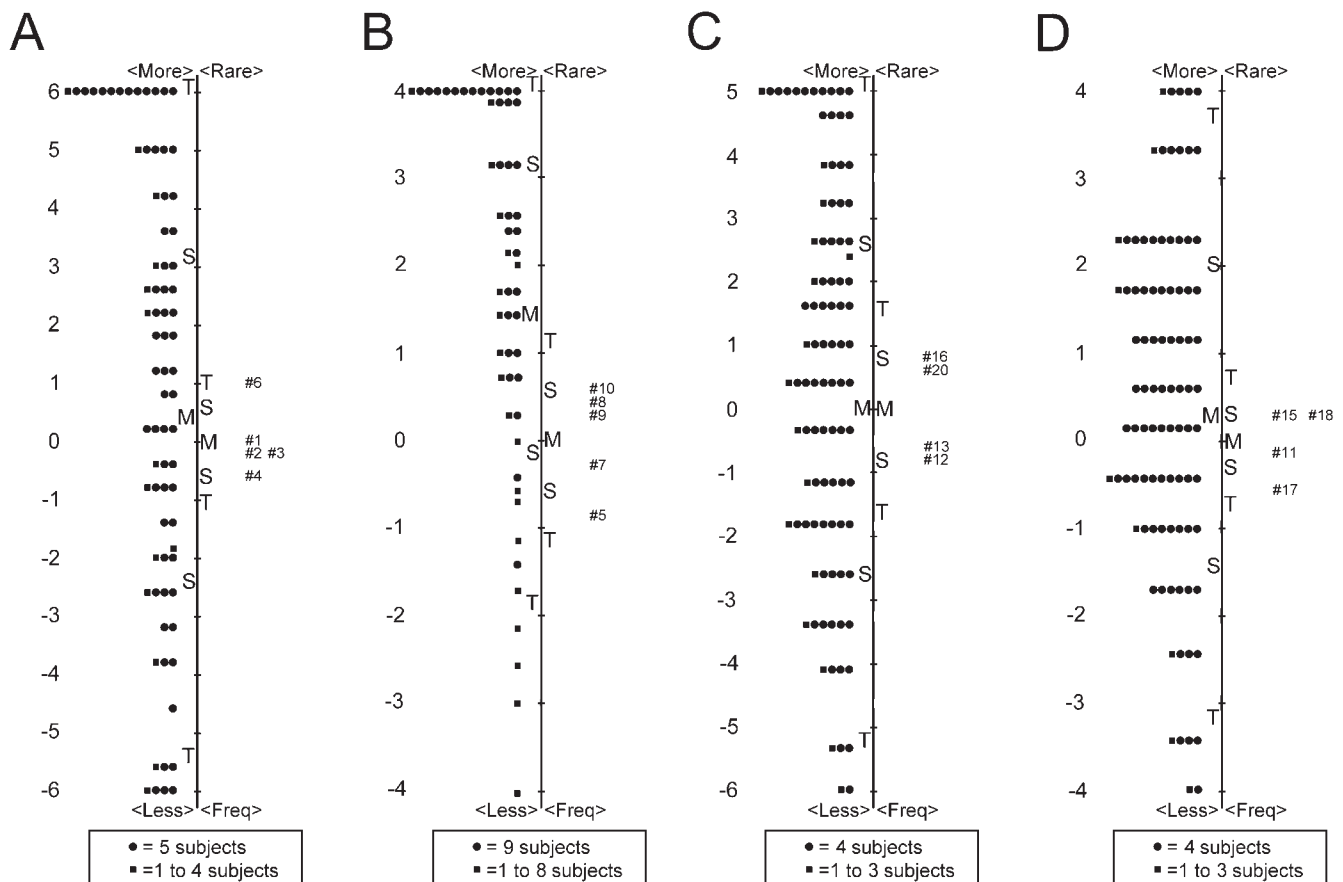


FIGURE 2. Adult Strabismus Questionnaire 20 (AS-20) targeting for (A) self-perception subscale, (B) interaction subscale, (C) reading function subscale, and (D) general function subscale. The mean severity discrimination (difficulty) of the items was well matched to the mean severity (ability) of the strabismus for the self-perception subscale (mean 1.07 ± 3.90 logits for person vs. 0.00 ± 0.51 logits for items), but targeting for the interaction subscale was relatively poor (mean 2.66 ± 2.26 logits for person vs. 0.00 ± 0.54 logits for items). Items were well matched to the patient cohort for the reading function subscale (mean 0.62 ± 3.23 logits for person vs. 0.00 ± 0.76 logits for items) and the general function subscale (mean 0.45 ± 2.04 logits for person vs. 0.00 ± 0.34 logits for items). Despite a relatively narrow range of item difficulty for each subscale, a wide range of severity of the condition was evident. M, Mean; S, 1 standard deviation; T, 2 standard deviations.

reliability ≥ 0.8 . In contrast, both the person separation index and reliability for the interaction and general function subscales are not optimal, limiting their ability to assess change in the underlying trait for an individual. Nevertheless, despite suboptimal person separation index and reliability, the interaction and general function subscales may be sufficient to assess the effect of strabismus on HRQOL in larger cohort studies, where group summary statistics are less sensitive to noise. We suggest that care should be taken when interpreting results of the interaction and general function subscales in individual patients, whereas the self-perception and reading function subscales appear to be robust.

Overall, the original AS-20 already met many of the conditions of the Rasch model, perhaps because of rigorous steps that were taken during the development of the questionnaire.¹ As previously reported, in the initial development of the AS-20, the original 181 patient-derived questions were administered to a cohort of 29 adult patients with strabismus. Items were eliminated if $\geq 10\%$ of responses were “not applicable” to their condition, $>80\%$ of patient responses were “never” or “rarely” (eliminating a ceiling effect), if an item received one or more negative comments, if the item was not applicable following treatment (e.g., surgery), if the item was discriminatory (socioeconomic, cultural, education bias), or if the item was more descriptive of strabismus symptoms rather than HRQOL. After applying these criteria, 49 items

remained and factor analysis was performed. The 10 highest loading items were retained in each of two factors, resulting in the original AS-20.¹

With only 20 items, the testing burden of the original AS-20 is reasonable and one could argue that removal of items based on failing to meet predefined statistical criteria alone may be inadvisable as long as the subscale being measured performs adequately. The two items considered for removal during our Rasch analysis were items 14, which pertains to problems with depth perception, and 19, which pertains to an inability to enjoy hobbies. Of note, some subjects express confusion over the meaning of depth perception in item 14. Likewise, some subjects express confusion on how to answer item 19, stating that they either do not have hobbies or they have multiple hobbies and have difficulty with some but not others. Given these comments from patients, it is not surprising that these two items show larger infit and outfit errors than other AS-20 items. Removal of these items from their respective subscales resulted in an improvement of subscale performance as measured by the person separation index and person reliability coefficient. Therefore, despite an already low burden of testing, the decision was made to remove these two items during Rasch analysis.

The range of item difficulty for the remaining items in each of the two psychosocial subscales and two function subscales was relatively narrow. Ideally, item difficulty would be more

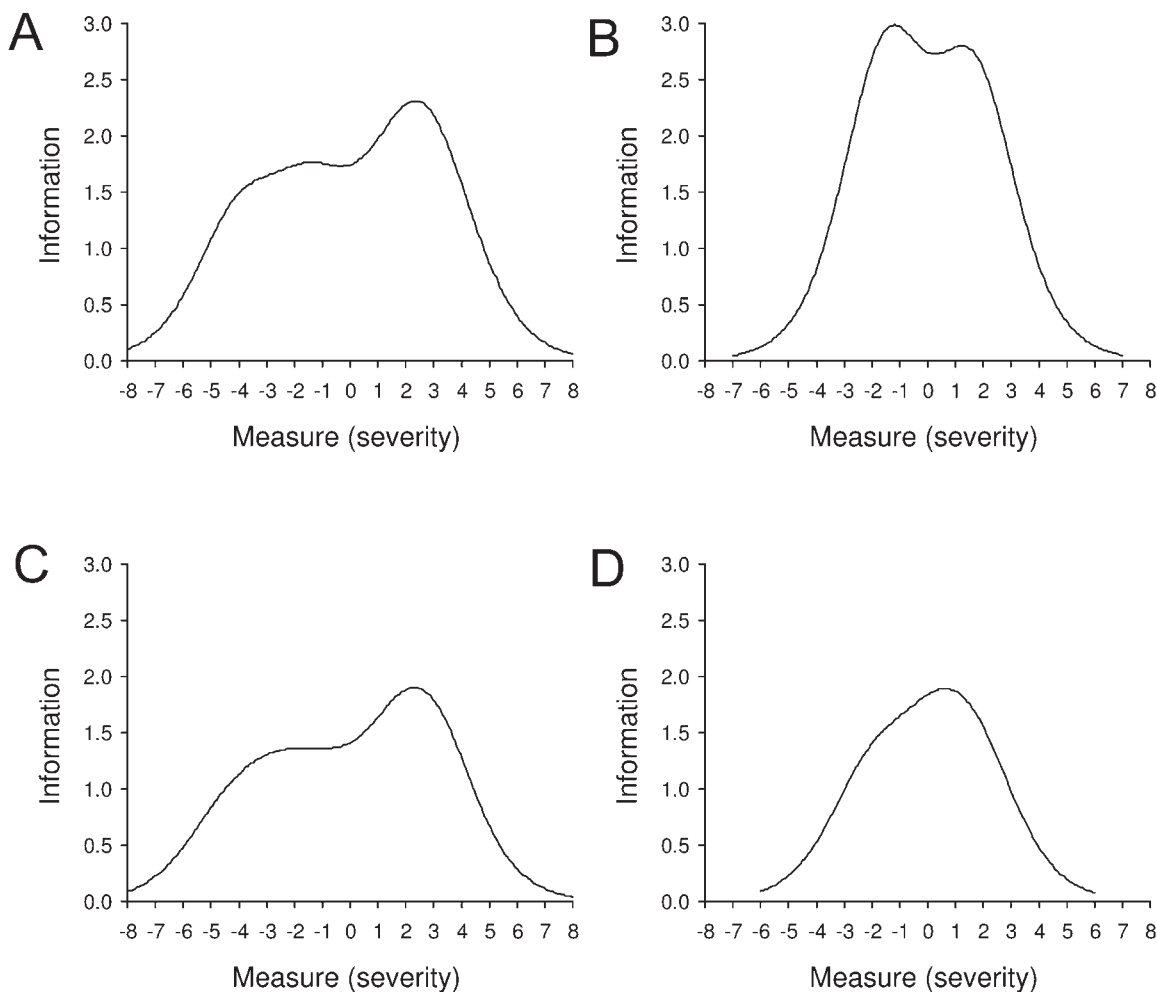


FIGURE 3. Adult Strabismus Questionnaire 20 (AS-20) test information curves following Rasch item reduction for (A) self-perception subscale, (B) interaction subscale, (C) reading function subscale, and (D) general function subscale.

widely distributed. Despite the relatively narrow range of item difficulty, the wide range of patient responses suggests items discriminate well between different severities of strabismus. This characteristic is potentially important when using the AS-20 to measure change in the severity of strabismus in response to treatment, such as surgery, as well as any treatment group comparisons in future studies.

Previous studies using non-Rasch scoring methods have reported an overall score for the AS-20 as well as individual scores for each of the subscales.^{3-5,16,17} Reporting composite AS-20 scores is problematic because there may be large offsetting changes in subscales that may result in a composite score that does not change in response to treatment. A similar problem exists for the originally described psychosocial and function subscales because we now report two subscales within each of these. Therefore, we now recommend that the AS-20 be Rasch-scored and reported as four separate subscale scores rather than a composite score. Similar recommendations have been made for the National Eye Institute Visual Function Questionnaire (NEI VFQ).¹¹ We have created conversion tools using commercial spreadsheet programs (Excel spreadsheets; Microsoft, Inc. Redmond, WA) to easily convert raw AS-20 responses to Rasch-scaled responses. These conversion tools are available online at www.pedig.org, by contacting the corresponding author, or online through the

journal as supplemental material (<http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8308/-DCSupplemental>).

So that Rasch person measures obtained from the conversion tools are more easily interpreted, these measures can be converted from a logit value to a 0 to 100 value (0 = worst HRQOL; 100 = best HRQOL) within the conversion tool through a linear transformation of the person scores. The minimum value for each subscale (value of 0) is determined by the average of each item's logit value when an "always" response is given for all items in that subscale, and the maximum (value of 100) when a "never" (or "never/rarely" for the general function subscale) is given for all items in the subscale. Missing responses may result in mean person measures above the maximum or below the minimum, so we therefore assign a "100" or "0" value, respectively. It is recognized that by assigning the maximum or minimum values on the 0 to 100 scale to subjects with missing values that the range of responses is technically narrowed, but the alternative is to assign a value that does not reach the extremes when a subject answers every single item with the maximum or minimum response (e.g., all responses are "never"). The addition of the 0 to 100 scoring option may facilitate application of Rasch scoring to existing and future data sets.

One of the main goals of Rasch analysis is to transform a traditionally scored instrument onto a linear scoring scale. Such transformation requires that raw scores be converted to a

TABLE 7. Dimensionality Analysis of the AS-20 Self-Perception, Interaction, Reading Function, and General Function Subscales

	Self-Perception Subscale		Interaction Subscale		Reading Function Subscale		General Function Subscale	
	Eigen	%	Eigen	%	Eigen	%	Eigen	%
Total raw variance	22.0	100.0	13.2	100.0	15.2	100.0	9.2	100.0
Explained by measures	17.0	77.2	8.2	62.0	11.2	73.7	5.2	56.6
Explained by persons	15.1	68.8	5.2	39.3	10.5	69.2	3.6	38.6
Explained by items	1.9	8.5	3.0	22.7	0.7	4.6	1.7	18.0
Total unexplained	5.0	22.8	5.0	38.0	4.0	26.3	4.0	43.4
1st contrast	1.8	8.1	1.7	12.8	1.6	10.4	1.6	17.7
2nd contrast	1.4	6.2	1.3	9.9	1.3	8.5	1.4	15.0
3rd contrast	1.0	4.3	1.0	7.8	1.1	7.3	1.0	10.7
4th contrast	0.9	4.1	1.0	7.3	0.0	0.1	0.0	0.0
5th contrast	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	Item	1st Contrast	Item	1st Contrast	Item	1st Contrast	Item	1st Contrast
	Number	Loading	Number	Loading	Number	Loading	Number	Loading
	1	-0.66	5	0.34	12	0.37	11	0.86
	2	0.26	7	0.52	13	0.79	15	0.19
	3	0.59	8	-0.66	16	-0.67	17	-0.66
	4	0.69	9	0.68	20	-0.61	18	-0.64
	6	-0.68	10	-0.64				

Rasch measure using a look-up table (such as those provided in this manuscript) or via mathematical equations, potentially limiting the ease of use among clinicians. To explore whether it is still reasonable (albeit not preferable) to use non-Rasch-scored AS-20 results for the 18 items remaining after Rasch analysis when informally assessing HRQOL in patients with strabismus, we plotted a test-characteristic curve for each of the four subscales showing the individual patient logit scores against the individual patient raw scores. Interestingly, these plots demonstrated a near-linear relationship for all subscales in all but the most extreme responses (data not shown), indicating that the traditional scoring system for each of the four AS-20 subscales (described in the present report) behaves in an essentially linear fashion. Additionally, individual threshold curves on the response item plot were uniformly spaced, again suggesting near linearity of the Likert-type response options within each item. We therefore propose an alternate scoring option using the traditional method of scoring the AS-20 as a reasonable and simple approach, based on the mean of all completed items for each of the four subscales. The self-perception (items 1, 2, 3, 4, and 6), interaction (items 5, 7, 8, 9, and 10), reading function (items 12, 13, 16, and 20), and general function (items 11, 15, 17, and 18) subscales may be scored with five responses ("Never" = 100, "Rarely" = 75, "Sometimes" = 50, "Often" = 25, and "Always" = 0). This proposed method of scoring the revised AS-20 is more convenient for clinicians than an alternative look-up table¹¹ to obtain Rasch-calibrated logit scores. Nevertheless, using the Rasch-scored AS-20 is preferable for any future research studies investigating the impact of strabismus on HRQOL.

The AS-20 benefits from analysis as four subscales (self-perception, interaction, reading function, and general function), removing items 14 and 19, and from reducing the response options for the general function subscale from five to four categories. The self-perception and reading function subscales had excellent person separation and reliability, whereas the interaction and general function subscales should be used with caution for individual patients due to suboptimal person separation and reliability. The Rasch-revised AS-20 may prove to be more useful than the original instrument.

References

- Hatt SR, Leske DA, Bradley EA, Cole SR, Holmes JM. Development of a quality-of-life questionnaire for adults with strabismus. *Ophthalmology*. 2009;116:139-144.
- Carlton J, Kaltenthaler E. Health-related quality of life measures (HRQoL) in patients with amblyopia and strabismus: a systematic review. *Br J Ophthalmol*. 2011;95:325-330.
- Hatt SR, Leske DA, Bradley EA, Cole SR, Holmes JM. Comparison of quality of life instruments in adults with strabismus. *Am J Ophthalmol*. 2009;148:558-562.
- Hatt SR, Leske DA, Holmes JM. Responsiveness of health-related quality of life questionnaires in adults undergoing strabismus surgery. *Ophthalmology*. 2010;117:2322-2328.
- Leske DA, Hatt SR, Holmes JM. Test-retest reliability of health-related quality of life questionnaires in adults with strabismus. *Am J Ophthalmol*. 2010;149:672-676.
- Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 2nd ed. Mahwah, NJ: Erlbaum; 2007:340.
- Pesudovs K. Patient-centred measurement in ophthalmology—a paradigm shift (Abstract). *BMC Ophthalmol*. 2006;6:25. Available at: <http://www.biomedcentral.com/content/pdf/1471-2415-6-25.pdf>. Accessed April 13, 2012.
- Brady CJ, Keay L, Villanti A, et al. Validation of a visual function and quality of life instrument in an urban Indian population with uncorrected refractive error using Rasch analysis. *Ophthalmic Epidemiol*. 2010;17:282-291.
- Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Measuring outcomes of cataract surgery using the Visual Function Index-14. *J Cataract Refract Surg*. 2010;36:1181-1188.
- Marella M, Pesudovs K, Keeffe JE, O'Connor PM, Rees G, Lamoureux EL. The psychometric validity of the NEI VFQ-25 for use in a low-vision population. *Invest Ophthalmol Vis Sci*. 2010;51:2878-2884.
- Pesudovs K, Gothwal VK, Wright T, Lamoureux EL. Remediate serious flaws in the National Eye Institute Visual Function Questionnaire. *J Cataract Refract Surg*. 2010;36:718-732.
- Vianya-Estopa M, Elliott DB, Barrett BT. An evaluation of the Amblyopia and Strabismus Questionnaire using Rasch analysis. *Invest Ophthalmol Vis Sci*. 2010;51:2496-2503.

13. Linacre JM. *Winsteps Rasch Measurement Computer Program User's Guide*. Beaverton, OR: Winsteps.com; 2011.
14. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci*. 2007;84:663-674.
15. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Rasch analysis of the quality of life and vision function questionnaire. *Optom Vis Sci*. 2009;86:E836-E844.
16. Durnian JM, Owen ME, Marsh IB. The psychosocial aspects of strabismus: correlation between the AS-20 and DAS59 quality-of-life questionnaires. *J AAPOS*. 2009;13:477-480.
17. Durnian JM, Owen ME, Baddon AC, Noonan CP, Marsh IB. The psychosocial effects of strabismus: effect of patient demographics on the AS-20 score. *J AAPOS*. 2010;14:469-471.