ARTICLE

Remediating serious flaws in the National Eye Institute Visual Function Questionnaire

Konrad Pesudovs, PhD, Vijaya K. Gothwal, PhD, Thomas Wright, Ecosse L. Lamoureux, PhD

PURPOSE: To test the assumption that the National Eye Institute Visual Function Questionnaire (NEI VFQ) measures visual functioning, assess the validity of its subscales, and, if flawed, revise the questionnaire and derive a shortened version with sound psychometric properties.

SETTING: Flinders Medical Centre, Adelaide, Australia.

METHODS: Patients from the cataract surgery waiting list self-administered and completed the 39-item NEI VFQ (NEI VFQ-39). Rasch analysis was applied, and the psychometric performance of the entire questionnaire and each subscale was tested. Instrument revision was performed in the context of Rasch analysis statistics.

RESULTS: Five hundred thirty-six patients (mean age 73.8 years) completed the questionnaire. Response categories for 2 question types were not used as intended so dysfunctional categories were combined. The NEI VFQ-39 and the 25-item version (NEI VFQ-25) had good precision but evidence of multidimensionality (more than 1 construct in 1 score), questions that did not fit the construct, suboptimum targeting of item difficulty to person ability, and dysfunctional subscales (8 NEI VFQ-39; 12 NEI VFQ-25). Questions could be reorganized into 2 constructs (a visual functioning scale and a socioemotional scale) that, after misfitting questions were removed, gave valid measurement of each construct and preserved 3 subscales. Removing redundancy from these long-form subscales yielded valid short-form scales.

CONCLUSIONS: Several NEI VFQ subscales were not psychometrically sound; as an overall measure, it is flawed by multidimensionality. This was repaired by segregation into visual functioning and socioemotional scales. Valid long and short forms of the scales could enhance application of the questionnaire.

Financial Disclosure: No author has a financial or proprietary interest in any material or method mentioned.

J Cataract Refract Surg 2010; 36:718–732 © 2010 ASCRS and ESCRS

Supplemental material available at www.jcrsjournal.org

The National Eye Institute Visual Function Questionnaire (NEI VFQ), one of the most commonly used patient-reported outcome measures in ophthalmic research,[1,2] was first described in 1998.[3] Since that time, the importance of measuring patient-reported outcomes has moved beyond question,[4–7] not only because numerous ophthalmology studies use patient-reported outcomes but also because patient-reported outcomes are now essential to U.S. Food and Drug Administration guidelines,[8] Medicare pays for performance initiatives depending on these reported outcomes,[9] and cost effectiveness studies that influence practice guidelines and 3rd-party payer policies use these outcomes.[10] Therefore, it is critical for ophthalmology to have high-quality patient-reported outcomes and for ophthalmologists to apply them appropriately.

What remains in question is the validity of individual patient-reported outcomes, in particular whether they measure what they purport to measure. According to traditional criteria, the NEI VFQ is ranked among the better patient-reported outcomes instruments with good construct validity.[11] An essential characteristic of a valid questionnaire is unidimensionality, meaning that all questions included in a single score contribute to the measurement of a single construct.[12] If unidimensionality is violated, it is not appropriate to report a single overall score derived from all questions—typically the main output of patient-reported outcomes such as

Published by Elsevier Inc.

0886-3350/10/$—see front matter
doi:10.1016/j.jcrs.2009.11.019

the NEI VFQ. To draw a clinical parallel, consider a new device that measures visual acuity and intraocular pressure (IOP) simultaneously and reports a single score from 0 to 100—a time-saving test perhaps, but useless for clinical care unless the results could be output as 2 individual (unidimensional) measures. The same is true for questionnaires in which a series of vision-related questions must be shown to measure the same construct. If they do not, they may be no more related than visual acuity is to IOP.

The NEI VFQ also reports scores on 12 subscales. This breadth of content encouraged the authors to suggest it measures vision-related quality of life, [3,13] a broad construct encompassing many effects on the person, such as well being, independence, and convenience. However, the name of the instrument suggests the construct under measurement is visual functioning (difficulty performing activities due to vision) rather than quality of life. It is possible that it measures both quality of life and visual functioning if the latter is a subset of the former. Although there is overlap between these 2 concepts, evidence suggests they are not interchangeable.[14] Recent studies[15–20] indicate that the NEI VFQ does not measure quality of life but does measure visual functioning. Therefore, the key questions are as follows: How many constructs does the NEI VFQ measure, and what do they represent?

We used Rasch analysis, a simple statistical transformation, to answer these questions. Rasch analysis transforms raw questionnaire data, which are composed of nominal numeric values applied to response options, into a continuous scale with interval-level measurement properties (like a ruler).[21–23] This transformation is important for scoring because it reduces noise and enables valid parametric statistical analysis of the output. Rasch analysis also gives unparalleled insight into the psychometric properties of questionnaires, including appropriateness of the response categories, measurement precision, unidimensionality, and item fit to the construct. These attributes have led to growing recognition of the value of Rasch analysis in the development or revalidation of questionnaires,[24] and the method is now used widely in ophthalmology.[25–31] As with other instruments,[20,32–35] it is appropriate to test the psychometric properties of the NEI VFQ using Rasch analysis. In particular, we used Rasch analysis because it would be useful for validating the unidimensionality of the NEI VFQ and for obtaining an interval-level measurement scale. Previous studies[19,20,36,37] assessed the interval-level properties of the NEI VFQ in a low-vision population, but only for a subset of visual functioning items. Thus, the studies provide limited insight into the performance of the NEI VFQ in its native form. Other studies evaluating NEI VFQ translations in a Rasch model did not fully evaluate its overall and subscale validity, including unidimensionality.[38,39] Assessment of the performance of an entire instrument does not confer validity to its subscales. Subscales have the same requirements for validity (unidimensionality, precision, item fit to the construct) as the overall instrument; therefore, it subscales must be individually tested.

The primary aim of this study was to explore the psychometric properties of the overall NEI VFQ and its 12 subscales using Rasch analysis to test the assumption that the underlying construct was visual functioning and whether other quality-of-life constructs were validly measured. Because this approach could expose deficiencies in the NEI VFQ, our second aim was to determine appropriate remedial measures to optimize the measurement properties of the NEI VFQ. A common barrier to the implementation of patient-reported outcomes is respondent burden,[24] resulting in numerous ad hoc attempts to shorten the NEI VFQ.[37,40] Therefore, the third aim of this study was to determine a way to shorten the NEI VFQ while retaining its sound psychometric properties, thus increasing the NEI VFQ's value to and use by the ophthalmic community. An additional aim was to facilitate interval-level scoring with the NEI VFQ for all users by providing ready-to-use spreadsheets that convert raw data to Rasch-scaled interval scores.

Submitted: August 20, 2009.
Final revision submitted: October 30, 2009.
Accepted: November 25, 2009.

From the NH&MRC Centre for Clinical Eye Research (Pesudovs, Gothwal, Wright), Department of Optometry and Vision Science, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, South Australia, Centre for Eye Research Australia (Lamoureux), Department of Ophthalmology, University of Melbourne, Victoria, and Vision CRC (Lamoureux), Sydney, Australia; Meera and LB Deshpande Centre for Sight Enhancement (Gothwal), Vision Rehabilitation Centres, L.V. Prasad Eye Institute, Hyderabad, India; Singapore Eye Research Institute (Lamoureux), Singapore National Eye Centre, Singapore.

Supported in part by National Health and Medical Research Council, Centre of Clinical Research Excellence Grant 264620 (Dr. Pesudovs), Canberra, Australia, and Career Development Awards 426765 (Dr. Pesudovs) and 529902 (Dr. Lamoureux).

Corresponding author: Konrad Pesudovs, NH&MRC Centre for Clinical Eye Research, Department of Optometry and Vision Science, Flinders Medical Centre, Bedford Park, South Australia, 5042, Australia. E-mail: konrad.pesudovs@flinders.edu.au.

MATERIALS AND METHODS

National Eye Institute Visual Function Questionnaire

The original NEI VFQ consisted of 51 questions (known as items in questionnaire research).[3,41] A shorter 26-item version (NEI VFQ-25)[13] with 13 additional optional items placed in an appendix (NEI VFQ-39) was developed. The NEI VFQ-39 has 39 items grouped into the following 12 vision-specific subscales: general health, general vision, ocular pain, near activities, distance activities, social functioning, mental health, role difficulties, dependency, driving, color vision, and peripheral vision. The NEI VFQ-25 consists of 26 items; 25 are used to compute the overall score, grouped into the same 12 vision-specific subscales, although 8 have fewer items.

Study Population

Since December 2005, as part of a long-term Cataract Outcomes Assessment Study, data on several cataract-specific questionnaires (including the NEI VFQ-39) have been collected. Packs of 10 questionnaires were routinely mailed to consecutive patients on the waiting list for cataract extraction surgery at Flinders Medical Centre, Adelaide, South Australia. Participants were asked to complete as many questionnaires as they could manage; they were given prepaid envelopes in which to return the questionnaires.

Ethical approval for the study was obtained, and all patients who agreed to participate signed a consent form. The study was conducted in accordance with the Declaration of Helsinki.

Patients were 18 years or older, English speaking, and had no severe cognitive impairment. The patients were representative of the elderly cataract population in Australia.[42]

Routine clinical assessments were performed. The evaluation included logMAR visual, which was reported binocularly because it is more representative of real-world ability.[43]

Rasch Analysis

The data analysis consisted of 3 phases. Phase 1 assessed the native instruments (25-item and 39-item versions). In phase 2, a minimalist reengineering approach was used to repair problems identified during the assessment phase. In phase 3, short-form versions of the reengineered NEI VFQ scales were developed.

Data for the analysis of the 39-item and 25-item versions of the NEI VFQ were extracted from the 39 items completed by participants. Data were recoded using the scoring instructions of the NEI VFQ (version 2000); the category ''stopped doing this for other reasons or not interested in doing this'' was treated as missing data for questions on difficulty. The response polarity was reversed for items 17 through 25 to render them consistent with other items, for which a higher score meant worse performance. The Winsteps program (version 3.67) was used for Rasch analysis by applying the Andrich rating scale model using joint maximum-likelihood estimation.[44] A group analysis with 1 rating scale model per question format (7) was used. This approach has been described.[36,45]

Rasch analysis locates item difficulty and person ability on a logit scale. A logit (log-odds unit) is the log odds ratio of the probability a person will endorse a particular rating scale step in an item over 1− the same probability. Therefore, persons of higher ability and items of greater difficulty are located on the negative side of the logit scale, while persons of lower ability and items of less difficulty are located on the positive side. Rasch analysis consists of the following components: category threshold order, person separation, unidimensionality, targeting, and differential item functioning (DIF).

Category Threshold Order The first step was to assess the ordering of the response category threshold. If one considers response options positioned along a scale, a threshold represents the transition between response options, which occurs when the likelihood of endorsing 1 category becomes the same as the likelihood of endorsing the next category. Most items on the NEI VFQ have 5 choices of increasing difficulty and, therefore, have 4 thresholds. Each threshold has a location on the logit scale, and each item has an average location. For each item, one would expect that with decreasing ability, the probability of selecting each statement in turn would increase in an ordered fashion from least to most difficult. Disordering of thresholds can result when a category is underused, its definition is not clear, or the number of categories exceed the number of levels the participants can distinguish.[46] Disordered thresholds can be a source of item misfit. Therefore, in cases of disordered thresholds, response categories were collapsed until thresholds were ordered; this was done before further analyses were performed.

Person Separation Person separation is a measure of precision and can be used to calculate how many groups or strata of person ability an instrument can discriminate.[21,47–49] The higher the person separation reliability, the more groups the instrument can define. A reliability coefficient of 0.8 indicates that 3 strata can be discriminated, and a reliability coefficient of 0.9 indicates 4 strata.[48] A person separation reliability of 0.8 is the minimum level of discrimination for an instrument to be considered satisfactory.[24]

Unidimensionality Rasch measurement requires unidimensionality of the instrument. Unidimensionality refers to whether the instrument measures a single underlying construct and whether each item ''fits'' the underlying construct.[50] Unidimensionality is assessed by examining the fit statistics and principal component analysis (PCA) of the residuals.[51] There are 2 types of item-fit statistics: infit and outfit. Both identify how well each item fits the construct; however, the infit statistic is less sensitive to distortion from outliers and is thus considered the more informative fit statistic.[46,50] Fit statistics were recorded as mean square standardized residuals (MNSQ); an infit MNSQ between 0.7 and 1.3 is considered acceptable.[52] Items with values less than 0.7 indicate a high level of predictability in the responses and thereby suggest redundancy. On the other hand, values higher than 1.3 indicate an unacceptable level of noise in the responses and are considered misfitting; ideally, they should be removed from the instrument.

The second test for unidimensionality examines the PCA of the residuals. A high level of variance accounted for by the principal component leads to a low likelihood of additional components; a variance of 60% or greater is considered good. Similarly, if the variance explained by the principal component for the empirical data and the variance for the model are comparable, the likelihood of additional constructs is low. The first contrast in the residuals reports whether there are patterns within variance that are unexplained by the principal component, which suggests a second construct is being measured. This study used the criterion that the contrast should have the strength of at least 2 items (measured by an eigenvalue >2.0) to be considered evidence of a second construct because this is greater than the magnitude seen with random data. The loading of items onto the contrasts allows identification of which items tap different constructs; in this study, a minimum loading of 0.4 was used to identify contrasting items.

Targeting Targeting can be visually inspected from the person-item map and refers to the extent to which the difficulty of the items matches the abilities of the persons in the sample. The person-item map also shows the item hierarchy and helps identify item gaps and redundant items (ie, items positioned at the same level of difficulty). Poor targeting occurs when items are clustered at certain points along the logit scale, leaving large gaps, and when many persons have a higher or lower ability than the most or least difficult item threshold. Targeting can also be measured by the difference between person and item mean values, which in a perfectly targeted instrument would be 0. A difference between means of more than 1 logit indicates notable mistargeting.

Differential Item Functioning An important characteristic of a good instrument is that items function similarly for persons at the same level of ability. Differential item functioning occurs when subgroups of people with comparable levels of ability respond differently to an item, which implies a response to some characteristic other than item difficulty.[50,53] For DIF testing, the respondents were stratified by sex, age (<74 years and ≥ 74 years), cataract status (bilateral versus awaiting cataract surgery in second eye), systemic comorbidity (present or absent), and ocular comorbidity (present or absent). Significance testing for DIF is sample size dependent[54]; therefore, in this study DIF was defined based on magnitude as follows: small or absent = difference less than 0.50 logit; minimal (but probably inconsequential) = difference 0.50 to 1.0 logit; notable = difference more than 1.0 logit.

Table 1. Sociodemographic and clinical characteristics of the 536 participants who completed the NEI VFQ.

Characteristic	Result
Mean age (y) ± SD	73.8 ± 9.6
Sex, n (%)	
Male	232 (43.3)
Female	304 (56.7)
Binocular visual acuity	
LogMAR	
Mean ± SD	0.22 ± 0.20
Range	−0.26 to 1.00
Snellen	
Mean	$6/9.5^{-1}$
Range	$6/3^{-2}$ to 6/60
Awaiting second-eye surgery, n (%)	220 (41.7)
Ocular comorbidity,* n (%)	
Present	256 (48.7)
Absent	270 (51.3)
Duration of cataract (y)	
Median	1
Interquartile range	3
Systemic comorbidity,† n (%)	
Present	427 (79.7)
Absent	64 (11.9)
Mean summary scores ± SD	
NEI VFQ-39	73.1 ± 19.1
NEI VFQ-25	71.8 ± 19.8

NEI VFQ-25 = 25-item National Eye Institute Visual Functioning Questionnaire; NEI VFQ-39 = 39-item National Eye Institute Visual Functioning Questionnaire

*Includes glaucoma, diabetic retinopathy, and age-related macular degeneration; data missing for 10 cases

†Includes diabetes, hypertension, and angina; data missing for 45 cases

Subscale Analysis

The 12 subscales were analyzed separately using the same rigorous procedures and criteria for reliability and validity that were used for the overall questionnaire. The subscales for color vision and peripheral vision contain only 1 item, which does not satisfy the criteria for Rasch measurement; therefore, these subscales were excluded from the analysis.

Statistical Analysis

Descriptive data were analyzed using SPSS software (version 15, SPSS Inc.). A P value less than 0.05 was considered statistically significant.

RESULTS

The pack including the NEI VFQ-39 was mailed to 1050 patients; 536 patients (51.0%) completed and returned the NEI VFQ-39. Table 1 shows the sociodemographic and clinical characteristics of the respondents and includes summary scores for the NEI VFQ-25 and NEI VFQ-39 to enable comparison of this population with others. Nonrespondents were demographically similar to the respondents. The mean age of the nonrespondents was 69.6 years; 55.9% were women, 44.1% had ocular comorbidity, and 76.5% had systemic comorbidity. The mean binocular visual acuity of nonrespondents was 0.31 logMAR ($6/12^{-}$ Snellen).

Phase 1: Assessment of Native Versions

Response Category Use On the NEI VFQ-39, category thresholds were ordered for all bar 2 question types. Figure 1 shows the first example of disordered category thresholds; at no point on the logit scale was the probability of responding to category 3 greater than the probability of responding to category 2 or 4. Therefore, this response category does not function as expected. Because category 3 is a neutral category, it did not seem logical to combine it with an adjacent category. Also, because only 7% of respondents chose this option, it was coded as a missing category. This repaired the disordered category thresholds. This question type was also present in the NEI VFQ-25. The other NEI VFQ-39 rating scale that contained 10

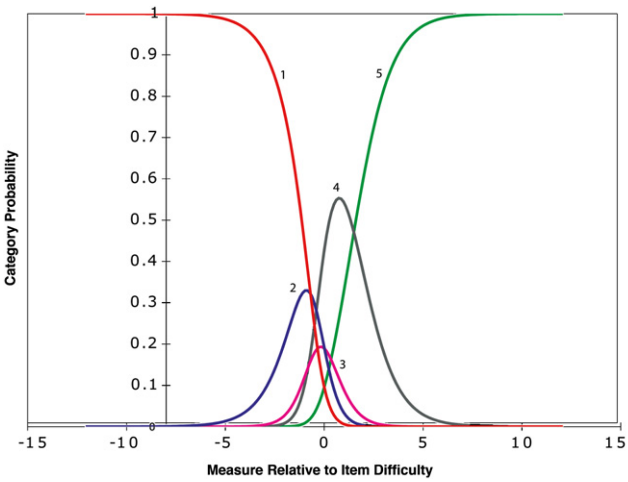

Figure 1. Category probability curves for the questions with agreement-type of response options in the NEI VFQ showing the range over which each of the 5 categories (1 = definitely true; 2 = mostly true; 3 = not sure; 4 = mostly false; 5 = definitely false) was likely chosen.

unlabeled categories showed disordered category thresholds (items A1 and A2). However, when category 5 was combined with category 4, reducing the 10 category response options to 9, the thresholds became ordered.

Overall Performance The person separation reliability values for both versions were excellent (0.9). Targeting was more than 1.0 logit indicating items were not ideally matched to persons in the sample (Table 2).

Item Fit and Dimensionality On the NEI VFQ-39, 5 items (12.8%) showed misfit, suggesting that the items did not measure the underlying construct and introduced noise into the measurement (Table 2 and Figure 2). The PCA of the residuals showed that the variance explained by the principal component was comparable for empirical calculation (60.6%) and by the model (60.7%). However, the unexplained variance explained by the first contrast was 3.3 eigenvalue units, the second contrast was 2.9 eigenvalue units, and the third contrast was 2.2 eigenvalue units; no further contrasts exceeded 2.0 eigenvalue units. This suggests that the questionnaire was not unidimensional. The principal component included all visual functioning items. Seven items loaded positively onto the first contrast and belonged to the mental health (1 item), role difficulties (2 items), and dependency (4 items) subscales. Three items loaded positively onto the second contrast and belonged to the social functioning (2 items) and color vision (1 item) subscales. One item loaded positively onto the third contrast and belonged to the mental health subscale. Similar results were found for NEI VFQ-25 (Table 2). However, the unexplained variance explained by the first contrast was 2.6 eigenvalue units, the second contrast was 2.1 eigenvalue units, and no further contrasts exceeded 2.0 eigenvalue units. Here, too, 5 items loaded positively onto the first contrast and belonged to the mental health (2 items) and dependency (3 items) subscales. Three items loaded positively onto the second contrast and belonged to the social functioning (2 items) and color vision (1 item) subscales.

Differential Item Functioning On the NEI VFQ-39, 7 items showed DIF by sex and presence of systemic comorbidity. Of the 7 items, 1 showed large DIF and the remaining showed minimal DIF. On the NEI VFQ-25, 3 items showed minimal DIF by sex and presence of systemic comorbidity (Table 3).

Table 2. Overall performance of all versions of the NEI VFQ.

	Phase 1		Phase 2				Phase 3	
Parameter	NEI VFQ-39	NEI-VFQ25	$LFVFS_{39}$	$LFSES_{39}$	$LFVFS_{25}$	$LFSES_{25}$	SFVFS	SFSES
Starting items (n)	39	25	23	16	14	11	15	12
Ending items (n)	39	25	15	12	8	10	6	7
Misfitting items (n)	5	4	0	0	0	0	0	0
Person separation reliability	0.96	0.93	0.91	0.88	0.87	0.87	0.86	0.84
Mean item location	0	0	0	0	0	0	0	0
Mean person location	−1.47	−1.52	−2.00	−1.94	−1.68	−1.64	−1.48	−1.60
Principal components analysis (eigenvalue in first contrast)	3.3	2.6	1.9	1.8	1.7	1.8	1.6	1.7
Valid subscales (n)	4	0	2	1	0	0	0	0

$LFSES_{25}$ = long-form socioemotional scale derived from NEI VFQ-25; $LFVFS_{25}$ = long-form visual functioning scale derived from NEI VFQ-25; $LFSES_{39}$ = long-form socioemotional scale derived from NEI VFQ-39; $LFVFS_{39}$ = long-form visual functioning scale derived from NEI VFQ-39; NEI VFQ-25 = 25-item National Eye Institute-Visual Functioning Questionnaire; NEI VFQ-39 = 39-item National Eye Institute-Visual Functioning Questionnaire; SFSES = short-form socioemotional scale; SFVFS = short-form visual functioning scale

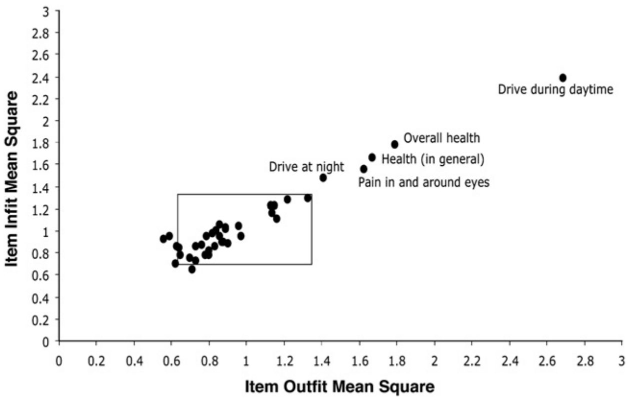

Figure 2. Scatterplot of infit versus outfit statistics for item measures estimated from responses to 39 items in the NEI VFQ-39. The box bounds the 0.7 to 1.3 acceptable limit; misfitting items lie outside the box.

Subscales Rasch analysis of the NEI VFQ-39 showed 4 of 10 subscales with satisfactory performance in person separation reliability (≥0.80 or greater), no misfitting items, and no multidimensionality (Table 4). The main problem with most subscales was insufficient person separation, showing the subscales lacked power to discriminate between individuals in the population. Two subscales (near activities and distance activities) showed some DIF, although it was minimal (magnitude ≤0.68 logits) and occurred for 5 items only. However, no subscale on the NEI VFQ-25 functioned satisfactorily.

Summary of Overall Performance Overall, the NEI VFQ-39 and NEI VFQ-25 had adequate discriminatory ability (good separation reliability); however, the targeting was not satisfactory, with most participants having better performance than the items allow. Moreover, there were significant numbers of misfitting items. Most important, the assumption of unidimensionality was violated. Furthermore, only 4 subscales on the NEI VFQ-39 and no subscale on the NEI VFQ-25 functioned satisfactorily.

With the absence of unidimensionality, it was unclear what the construct under measurement was. It was partly visual functioning and partly another construct(s). For the NEI VFQ to effectively measure patient-reported outcomes, it must be segregated into unidimensional constructs. The PCA showed that 3 constructs might be present. The first construct contained visual functioning items; the second, dependency, mental health, and role difficulty items; and the third, social functioning items. However, subscale analysis showed that the dependency, mental health, and social functioning items were all dysfunctional, lacking sufficient person separation. However, the PCA results suggest that some social and emotional subscales could be combined to provide sufficient items to achieve satisfactory person separation. Therefore, to enable measurement of constructs other than visual functioning, items from the contrasts (dependency, mental health, role difficulties, and social functioning) were combined. Next, it was determined whether unidimensional measurement was possible. Although there was no existing term to collectively represent the latter construct, it could be described as a socioemotional construct and is referred to as such hereafter. Therefore, to effectively measure patient-reported outcomes with the NEI VFQ, an evaluation was performed to determine whether the instrument could be reengineered around the 2 constructs (ie, visual functioning scale and socioemotional scale). It

Table 3. Items showing DIF on the NEI VFQ-39 and the NEI VFQ-25. The groups shown rated the item as relatively easier than the other items.

	Demographic Variable				
Item Description	Sex	Age	Cataract Status	Ocular Comorbidity	Systemic Comorbidity
Health (in general)		—	—	—	Without* (1.18)
Picking out and matching your own clothes	Female* (0.54) Female† (0.58)	—	—	—	With† (0.74)
Driving at night	Male* (0.72) Male† (0.68)	—	—	—	—
Driving in difficult conditions, such as in bad weather, during rush hour, on the freeway, or in city traffic	Male* (0.57) Male† (0.56)	—	—	—	—
Overall health	—	—	—	—	Without* (0.90)
Shaving, styling your hair, or putting on make up	Male* (0.69)	—	—	—	—
Recognizing people you know from across a room	—	—	—	—	With* (0.50)

*NEI VFQ-39
†NEI VFQ-25

Table 4. Overall performance of subscales on the NEI VFQ-39.

	Subscale									
Parameter	General Health*	General Vision	Ocular Pain	Near Activities*	Distance Activities*	Social Functioning	Mental Health	Role Difficulties*	Dependency	Driving
Items (n)	2	2	2	6	6	3	5	4	4	3
Misfitting items (n)	0	0	0	0	0	1	1	0	0	1
Person separation reliability	0.86	0.79	0.63	0.86	0.81	0.65	0.71	0.80	0.76	0.58
Mean item location	0	0	0	0	0	0	0	0	0	0
Mean person location	−2.03	−0.58	−2.98	−1.83	−1.89	−3.05	−0.77	−1.66	−1.46	−1.69
Principal components analysis† (eigenvalue)	0.2	—	—	1.5	1.2	—	—	1.9	—	—

*Satisfactorily functioning subscale
†Values of optimally functioning subscales only

was hypothesized that in addition to this most important modification, reengineering the scales would address the other problems (item misfit, targeting, DIF).

Phase 2: Reengineering the NEI VFQ

The second phase consisted of reengineering the NEI VFQ by making the minimum changes required to establish satisfactory measurement properties. Two separate scales were formed: the visual functioning scale and the socioemotional scale. Assuming minimum revision, the resulting scales were referred to as long-form scales. This applies to the NEI VFQ-39 and the NEI VFQ-25; for the sake of clarity, regardless of the number of items in the revised scales, the revised scales will be identified by their origin as [scale name]$_{39}$ or [scale name]$_{25}$.

Long-Form Visual Functioning Scale$_{39}$ The first revision of the NEI VFQ-39 visual functioning scale was the long-form visual functioning scale$_{39}$ (LFVFS$_{39}$). Following are the details:

1. Item fit and dimensionality Twenty-three items of the NEI VFQ-39 loaded onto the visual functioning dimension, although several items misfit. The first modification was to remove misfitting items. This was an iterative process starting with the most misfitting item. Eight items were deleted before all remaining items (15) fit the Rasch model (Tables 2 and 5). There was minimal loss of real person separation and slight worsening (0.5 logits) of targeting (Figure 3). The PCA of the residuals showed that the LFVFS$_{39}$ was unidimensional; the variance explained by the principal component was comparable for the empirical calculation (66.8%) and by the model (65.9%), and the unexplained variance explained by the first contrast was 1.9 eigenvalue units.

2. Differential item functioning Three (20%) items showed minimal DIF by demographic variables (ie, sex, age, ocular comorbidity). The most problematic item was "picking out and matching clothes," which was rated relatively easier by women (0.89 logits). The other 2 items showed DIF of less than 0.80 logits.

3. Subscales After item reduction, 2 valid subscales from the initial NEI VFQ-39 (near activities and distance activities) were included in the LFVFS$_{39}$ in their entirety. The near activities subscale had a person separation reliability of 0.86 and the distance activities subscale, of 0.81. Neither had misfitting items, and both were unidimensional by PCA, with first-contrast magnitudes of 1.5 eigenvalue units and 1.2 eigenvalue units, respectively.

Long-Form Socioemotional Scale$_{39}$ The first revision of the NEI VFQ-39 socioemotional scale was the long-form socioemotional scale$_{39}$ (LFSES$_{39}$). Following are the details:

1. Item fit and dimensionality Sixteen items from the NEI VFQ-39 were included in the socioemotional scale, although several items misfit. The first modification was to remove misfitting items. Four items were deleted iteratively before all remaining items (12) fit the Rasch model (Tables 2 and 5). There was minimal loss of person separation and slight worsening (0.4 logits) of targeting (Figure 4). The PCA of the residuals showed that the LFSES$_{39}$ was unidimensional; the variance explained by the measures was comparable for the empirical calculation (70.3%) and by the model (69.4%), and the unexplained variance explained by the first contrast was 1.8 eigenvalue units. There was no DIF.

2. Subscales After item reduction, 1 valid subscale from the initial NEI VFQ-39 (role difficulties) was included in the LFSES$_{39}$ in its entirety. Person

Table 5. Summary of items included in each revised version of the NEI VFQ.

		NEI VFQ-39/NEI VFQ-25 Version*					
Item	Description	$LFVFS_{39}$	$LFSES_{39}$	$LFVFS_{25}$	$LFSES_{25}$	SFVFS	SFSES
1	Health	–	–	–	–	–	–
2	Eyesight	+	–	+	–	+	–
3	Worry	–	–	–	–	–	–
4	Pain or discomfort	–	–	–	–	–	–
5	Read ordinary print in newspapers	+ (near)	–	+	–	+	–
6	See well up close	+ (near)	–	+	–	+	–
7	Find something on a crowded shelf	+ (near)	–	+	–	+	–
8	Read street signs or the names of stores	+ (distance)	–	+	–	+	–
9	Going down steps, stairs, or curbs in dim light or at night	+ (distance)	–	+	–	+	–
10	Notice objects off to the side while walking	+	–	+	–	–	–
11	See how people react to things	–	+	–	+	–	–
12	Pick and match own clothes	+	–	–	–	–	–
13	Visiting with people in their homes, or at parties, or in restaurants	–	+	–	+	–	+
14	Go out to see movies, plays, or sports events	+ (distance)	–	+	–	–	–
15c	Drive during daytime in familiar places	–	–	–	–	–	–
16	Drive at night	–	–	–	–	–	–
16a	Drive in difficult conditions	–	–	–	–	–	–
17	Accomplish less	–	+ (role)	–	+	–	+
18	Limited	–	+ (role)	–	+	–	+
19	Pain around eyes	–	–	–	–	–	–
20	Stay home most of the time	–	+	–	+	–	+
21	Frustrated	–	–	–	+	–	–
22	Much less control	–	+	–	+	–	+
23	Rely too much on what other people tell	–	+	–	+	–	+
24	Need a lot of help	–	+	–	+	–	–
25	Do things that will embarrass	–	+	–	+	–	+
A1	Overall health	–	–	–	–	–	–
A2	Eyesight now	–	–	–	–	–	–
A3	Read small print in a telephone book, on a medicine bottle, or on legal forms	+ (near)	–	–	–	–	–
A4	Figure out whether bills received are accurate	+ (near)	–	–	–	–	–
A5	Doing things like shaving, styling hair, or putting on makeup	+ (near)	–	–	–	–	–
A6	Recognize people across a room	+ (distance)	–	–	–	–	–
A7	Take part in active sports or other outdoor activities	+ (distance)	–	–	–	–	–
A8	See and enjoy programs on TV	+ (distance)	–	–	–	–	–
A9	Entertain friends and family in home	–	–	–	–	–	–
A11a	Have more help	–	+ (role)	–	–	–	–
A11b	Limited	–	+ (role)	–	–	–	–
A12	Irritable	–	–	–	–	–	–
A13	Don't go out of home alone	–	+	–	–	–	–

+ = item used in version; – = item not used in version; $LFSES_{25}$ = long-form socioemotional scale derived from NEI VFQ-25; $LFVFS_{25}$ = long-form visual functioning scale derived from NEI VFQ-25; $LFSES_{39}$ = long-form socioemotional scale derived from NEI VFQ-39; $LFVFS_{39}$ = long-form visual functioning scale derived from NEI VFQ-39; NEI VFQ-25 = 25-item National Eye Institute-Visual Functioning Questionnaire; NEI VFQ-39 = 39-item National Eye Institute-Visual Functioning Questionnaire; SFSES = short-form socioemotional scale; SFVFS = short-form visual functioning scale
*Items parentheses indicate those from valid subscales (near activities, distance activities, role difficulties) on the $LFVFS_{39}$ and $LFSES_{39}$.

separation reliability was 0.80, and the subscale was unidimensional.

Long-Form Visual Functioning Scale$_{25}$ The first revision of the NEI VFQ-25 visual functioning scale was the long-form visual functioning scale$_{25}$ ($LFVFS_{25}$). The results were similar to those for the revision of the respective NEI VFQ-39 scale. Fourteen items from the NEI VFQ-25 loaded onto the visual functioning construct, although several misfit. Six items were

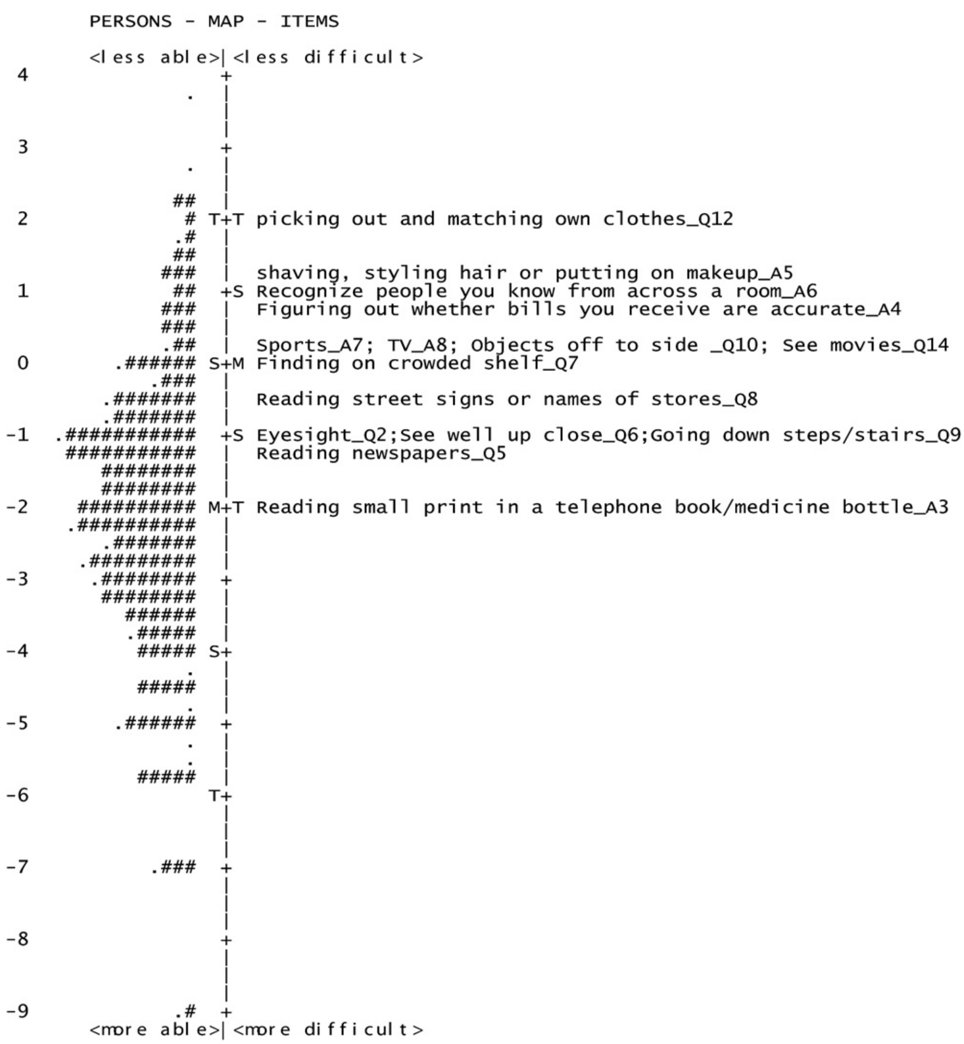

Figure 3. Person-item map of the $LFVFS_{39}$. The participants are on the left of the dashed line, with more able participants located at the bottom of the map. Items are located on the right of the dashed line, with more difficult items located at the bottom of the map. Each # represents 3 participants (M = mean; S = 1 SD from the mean; T = 2 SD from the mean).

deleted before all remaining items (8) fit the Rasch model (Table 2). The PCA of the residuals showed that the variance explained by the principal component was comparable for the empirical calculation (67.5%) and the model (67.1%). The first-contrast magnitude was 1.7 eigenvalue units.

Two items showed minimal DIF; older respondents rated "reading street signs or the names of stores" 0.61 logits easier than other items compared with younger respondents, and participants with systemic comorbidity rated "going out to see movies, plays, or sports events" 0.51 logits easier than other items compared with respondents with systemic comorbidity. No subscale was valid.

Long-Form Socioemotional Scale$_{25}$ The first revision of the NEI VFQ-25 socioemotional scale was the long-form socioemotional scale$_{25}$ ($LFSES_{25}$). The results were similar to those for the revision of the respective NEI VFQ-39 scale. Eleven items entered the analysis; 1 was deleted before all remaining items (10) fit the Rasch model (Table 2). The PCA of the residuals showed that the variance explained by the principal component was comparable for the empirical calculation (71.9%) and by the model (70.9%). The unexplained variance explained by the first contrast was 1.8 eigenvalue units. There was no DIF. No subscale was valid.

Summary The long-form scales built on 2 constructs (visual functioning and socioemotional) of the NEI VFQ-39 and NEI VFQ-25 showed good psychometric properties. Except for the general health subscale, the other 3 valid subscales were retained in their entirety; that is, near activities and distance activities on the $LFVFS_{39}$ and role difficulties on the $LFSES_{39}$. There were no functioning subscales in the long-form scales derived from the NEI VFQ-25 (ie, $LFVFS_{25}$ and $LFSES_{25}$). Despite being shorter, there was still redundancy in the content of the long-form instruments. Therefore, if one were willing to forego the presence of any subscales to pursue efficient measurement

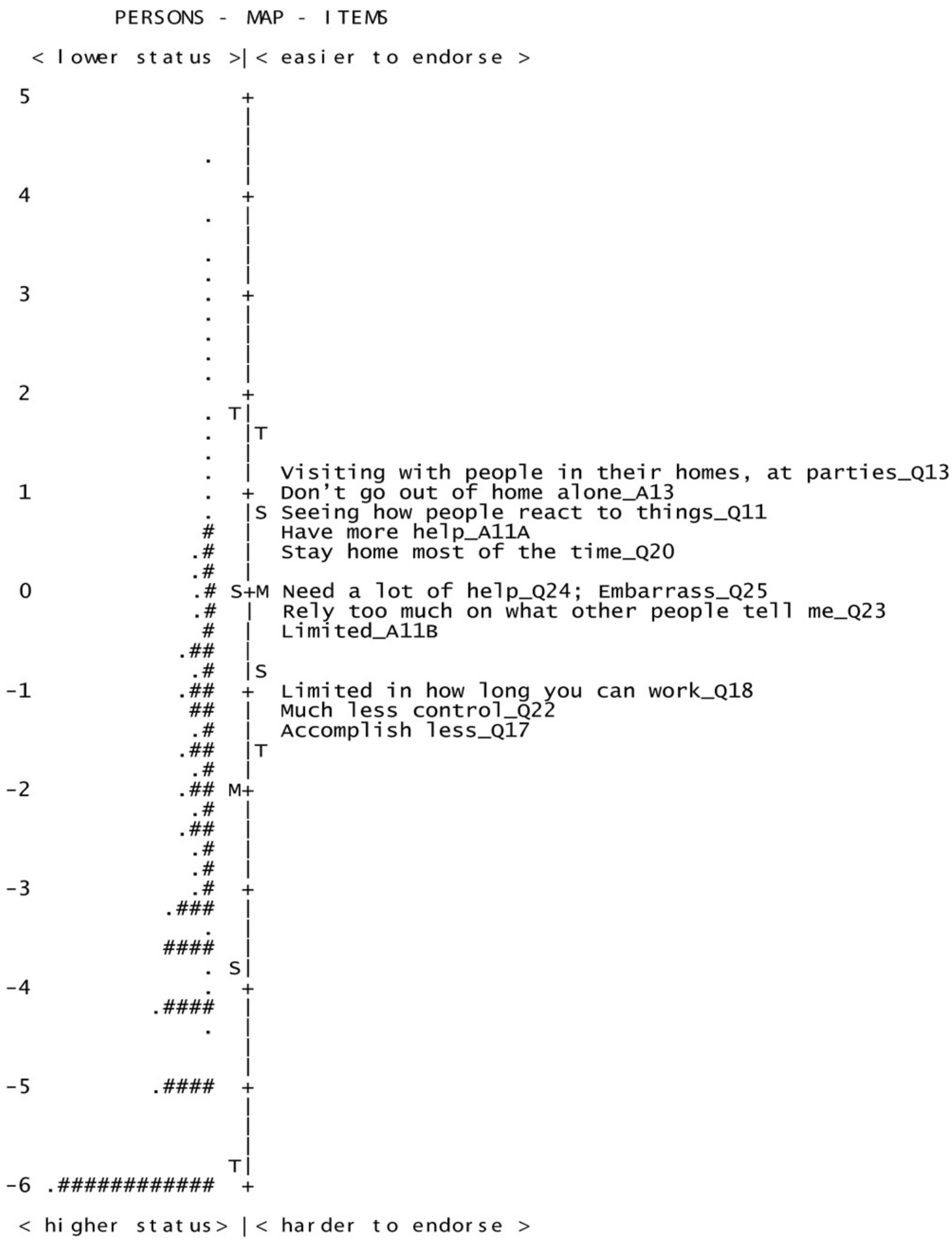

Figure 4. Person-item map of the $LFSES_{39}$. The participants are on the left of the dashed line, with the more able participants located at the bottom of the map. Items are located on the right of the dashed line, with more difficult items located at the bottom of the map. Each # represents 8 participants (M = mean; S = 1 SD from the mean; T = 2 SD from the mean).

with the shortest possible instruments, further revision was possible.

Phase 3: Developing Short-Form Versions

The following criteria were used to drive item removal: (1) Delete the most redundant item (lowest infit) and (2) delete the item with the poorest targeting but (3) maintain a minimum person separation reliability of 0.83. (Although the minimum acceptable was 0.80, using 0.83 as a limit would allow poorer performance in another population.)

An attempt was made to shorten the visual functioning and socioemotional scales derived from the 39-item and 25-item versions of the NEI VFQ; it was found that the short forms essentially converged to the same minimum item set. Therefore, only 1 short-form visual functioning scale (SFVFS) and 1 short-form socioemotional scale (SFSES) are reported.

In the 2 short-form versions, 2 items were removed from the $LFVFS_{25}$ and 3 items from the $LFSES_{25}$. The SFVFS contained 6 items with acceptable person separation reliability and was unidimensional (Table 2). The targeting (–1.48 logits) was better than that of the longer versions but was suboptimum (Figure 5). Two items showed minimal DIF by age and systemic comorbidity; that is, compared with younger respondents, older respondents rated "finding something on a crowded shelf" 0.58 logits easier than other items and respondents with systemic comorbidity rated "reading street signs or the names of stores" 0.54 logits easier than other items.

The SFSES contained 7 items with minimally acceptable person separation reliability and was unidimensional (Table 2). The targeting (–1.60 logits) was marginally better than that of the longer versions but was again suboptimum (Figure 6). There was no DIF. Although there was no redundancy, there were significant gaps between items. For example, there was a 0.93-logit gap between "limited in work" and "rely too much on what other people tell."

DISCUSSION

The first goal of this study was to test the unidimensionality of the NEI VFQ; unidimensionality was rejected. Although most items on the NEI VFQ tap the construct of visual functioning, other items belong to a different construct, namely socioemotional issues. Therefore, the NEI VFQ is not simply a measure of visual functioning. However, if the visual functioning items were segregated into a scale, it is a very effective measure of visual functioning. This is consistent with the work of Massof[18,19] and Massof and Ahmadian,[20] who showed that a subset of NEI VFQ functioning items taps the same construct as other visual functioning instruments; that is, the Visual Functioning 14 (VF-14), the Visual Activities Questionnaire, and the Activities of Daily Vision Scale (ADVS). Many other studies[38,39,55] applied Rasch analysis to the NEI VFQ, but simply to derive Rasch estimates of total and subscale scores. No attempt was made to assess the validity of the NEI VFQ within the Rasch model. This is an important distinction; that is, the use of Rasch analysis for scoring an instrument and its use as part of a comprehensive assessment of an instrument's psychometric properties are not the same. We believe our study is the first published attempt at the latter. Ryan et al.[37] used Rasch analysis to show valid measurement of functioning with a 7-item subset of the NEI VFQ. Although Rasch analysis was applied correctly and the problem of multidimensionality was avoided by dealing only with functioning items, the authors' choice of items was not based on statistical evidence. Rather, they chose the items most responsive to low-vision intervention. Therefore, Ryan et al.'s version is different from any of the visual functioning scales we suggest because our choice of items was based on optimizing the inherent psychometric properties of the NEI VFQ.

The problem with multidimensionality is that it invalidates the reporting of a total score derived from all items. This is because such a total score does not inherently represent a concept. For the NEI VFQ, the total score is a combination of visual functioning and

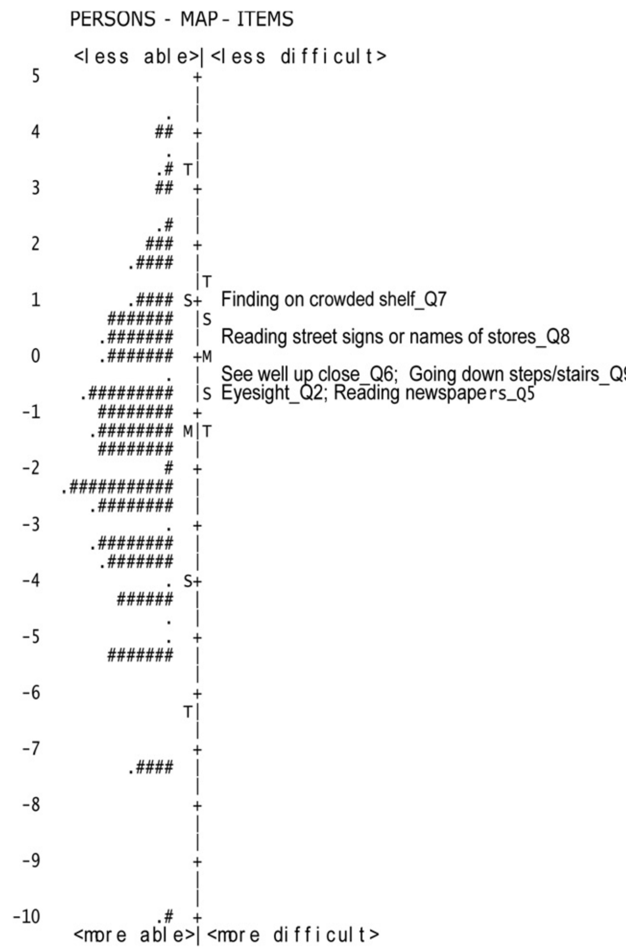

Figure 5. Person-item map of the SFVFS. The participants are on the left of the dashed line, with more able participants located at the bottom of the map. Items are located on the right of the dashed line, with more difficult items located at the bottom of the map. Each # represents 4 participants (M = mean; S = 1 SD from the mean; T = 2 SD from the mean).

socioemotional constructs. Although the 2 constructs are related, they are too different to be combined. One could argue that this represents vision-related quality of life. Even if this were so, it is no justification for reporting a total score that is really only a measure of visual functioning contaminated by the noise associated with social and emotional issues. Perhaps this is why the total score on the NEI VFQ-25 correlates poorly with a global rating of quality of life or health-related quality of life.[15,56,57] One can measure both socioemotional issues and visual functioning with the NEI VFQ; by reporting scores from the 2 dimensions, one can claim that vision-related quality of life is measured. However, it is likely other constructs make up the universe of vision-related quality of life.

The NEI VFQ was developed to have 12 subscales, which implies there are 12 measurable constructs. However, only 4 subscales met the criteria for valid

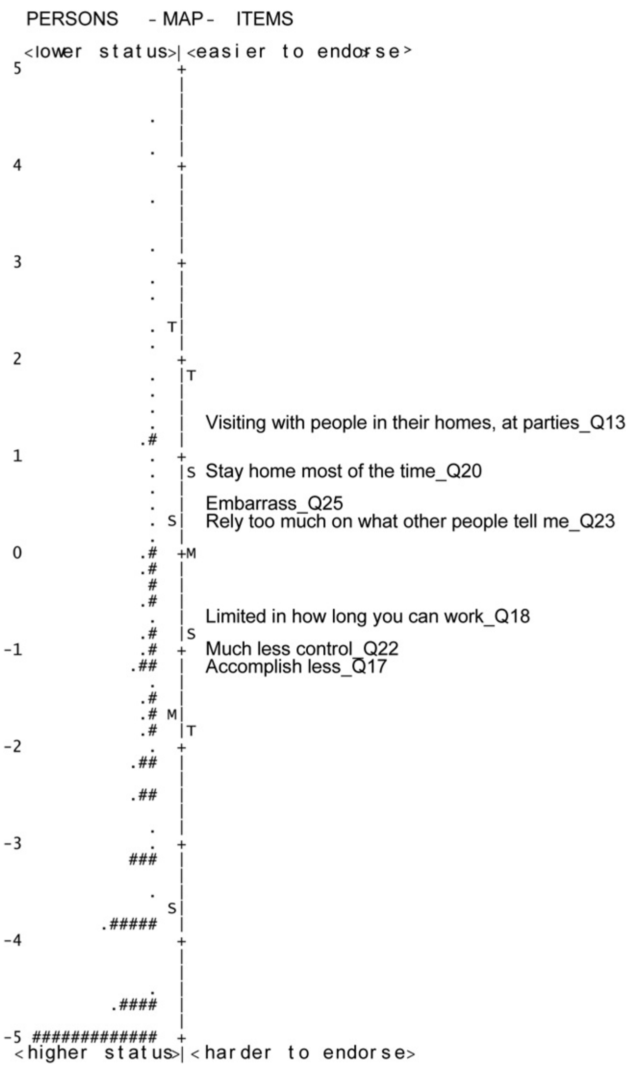

Figure 6. Person-item map of the SFSES. The participants are on the left of the dashed line, with more able participants located at the bottom of the map. Items are located on the right of the dashed line, with more difficult items located at the bottom of the map. Each # represents 9 participants (M = mean; S = 1 SD from the mean; T = 2 SD from the mean).

measurement. These were general health, near activities, distance activities, and role difficulties. The failure of the social functioning, mental health, and dependency subscales to measure effectively prompted us to explore whether combining these items would measure a socioemotional construct. Analyses showed these items tapped a single measurable construct, thus ensuring the retention of some measurement from these dysfunctional subscales. The main problem for the validity of the subscales was their inability to discriminate among the population under measurement. Fundamentally, this was the result of too few items, which with a limited number of response options cannot discriminate between people of a similar amount of construct. This problem is analogous to using a 1 m long ruler to measure height; all this will do is separate people into those more or less than 1 m tall and those taller than 2 m. However, by adding centimeter marks to the ruler, height can be measured much more precisely. Similarly, more items would have to be added to the subscales to provide discrimination. An appendix of new items could be added to the NEI VFQ to enable measurement of these subscales.

The addition of new items could also help resolve the problem of poor targeting. Whether in its original form or shortened forms, the NEI VFQ items poorly target our cataract population. This is shown in Figures 3 and 4, in which few items target the more able people. Returning to our meter ruler analogy, this indicates that even the reengineered NEI VFQ does not have gradations in all the necessary parts of the scale to discriminate all people. Therefore, simply adding more items to facilitate good measurement is not enough; adding appropriately targeted items (to get the marks in the right part of the ruler) is also required. These new items would probably have to be sought from patient focus groups, although it is possible that some items from the 51-item field test version of the NEI VFQ could help.[41] Poor targeting may also be sample dependent. Our cataract patients may not be as visually impaired as people with other eye diseases for whom targeting may be more appropriate. Although our NEI VFQ summary scores are comparable to those in other cataract surgery populations,[2] lower scores have been reported for eye diseases, such as diabetic macular edema (approximately 9 points lower) and age-related macular degeneration (approximately 2 points lower).[57,58] This does not mean that our respondents did not have visual disability but rather that the activities represented in the NEI VFQ were too easy for them. All the participants were drawn from a cataract surgical waiting list; thus, by definition, they had visual disability resulting from cataract,[59–61] as previously reported.[42] Suboptimum targeting is common in cataract patients and has been reported using other questionnaires, such as the ADVS[34] and VF-14.[31]

The NEI VFQ has been the most commonly used patient-reported outcome measure in ophthalmology over the past decade. This popularity is likely to persist, so we reengineered the NEI VFQ to optimize its measurement properties. Based on our study, 3 versions of the NEI VFQ can be recommended for use: long-form versions derived from the NEI VFQ-39 and the NEI VFQ-25 and a short-form version. All versions measure the same constructs: a visual functioning scale and a socioemotional scale. The long-form version derived from the NEI VFQ-39 has the advantage of retaining 3 subscales (near activities, distance activities, and role difficulties). This enables the reporting of 5 patient-reported outcome measures but has the disadvantage of being 27 items long. The long form derived from the NEI VFQ-25 has no subscales but is considerably shorter at 18 items. Short instruments enhance utility; therefore, we created a short-form version with 13 items that still reports visual functioning and socioemotional scores. The only disadvantage of the short-form version is reduced person separation reliability. Although the person separation reliability is satisfactory, better separation reliability means better measurement. Therefore, when possible, the long form derived from the NEI VFQ-39 should be used. If one were to reanalyze existing NEI VFQ data, one would use the long-form version of the NEI VFQ-39 when available or the long-form NEI VFQ-25 when only these data were collected.

Furthermore, to eliminate the need for researchers to perform Rasch analysis, we created ready-to-use Excel spreadsheets (Microsoft, Inc.) for conversion of raw scores to Rasch measurement for all versions of the NEI VFQ (available from the corresponding author or as Supplements A to F, available at www.jcrsjournal.org). These conversions assume the results in our study are generalizable to other populations. Although this may be true in some cases, item calibrations can vary across populations, albeit with much less error than simple application of ordinal values.[33] Targeting, and to a lesser extent DIF, are also likely to vary with populations. Therefore, it is always better to perform Rasch analysis of each data set than to rely on a conversion algorithm. However, the multidimensionality of the NEI VFQ is an inherent feature of the instrument that would be consistently demonstrable across populations.[62] Thus, identification of the flaws in the original instrument and the revisions we propose represent important progress for the use of the NEI VFQ. Nevertheless, our model should be tested in other populations.

The use of Rasch analysis is critical to the evaluation of visual functioning instruments because it enables thorough examination of the psychometric properties of the instrument and creates interval scoring. However, other item response theory approaches could be used. In particular, multilevel models allow simultaneous analysis of different dimensions in a multidimensional instrument.[63] However, segregation of the constructs and individual Rasch analyses is fundamentally a stepwise version of the same procedure. Of the various item response theory approaches, Rasch analysis remains favored for estimates of person and item parameters because the model contains no other unknowns, making the parameter estimates valid as measures. For this reason, Rasch analysis is becoming the gold standard for scoring patient-reported outcome measures in ophthalmology.

The reengineered versions of the NEI VFQ are not without flaw. The main problem is poor targeting, with a mistargeting greater than 1.0 logit being suboptimum. To further refine the NEI VFQ, items that target more able participants should be added. Although this could improve targeting, a better approach is to use Rasch analysis in the initial development of the instrument; there are many such examples in ophthalmology.[27,28,64,65] However, a better strategy would be to develop an item bank and computer-adaptive testing. In computer-adaptive testing, Rasch-calibrated items are presented based on the response to the previous item. This tailored approach has been used in other areas of health care and could be developed for ophthalmology. This strategy could lead to more precise and more accurate measurement across a greater range of traits for all patient-reported outcomes in ophthalmology. This could have far-reaching benefits for clinical research; for example, more precise measurement reduces the sample size needed for clinical trials and a greater range of trait allows application in more diverse populations with valid comparison of results.

In summary, the present study does not support the assumption that the NEI VFQ measures only visual functioning. Rather, valid measurement of visual functioning and a socioemotional construct is possible by using 1 of the reengineered versions of the NEI VFQ. Because most robust parametric statistical techniques assume that the data are at least on an interval scale, we recommend the Rasch-transformed scores of the NEI VFQ for future applications.

REFERENCES

1. Lin I-C, Wang I-J, Lei M-S, Lin LL-K, Hu F- R. Improvements in vision-related quality of life with AcrySof IQ SN60WF aspherical intraocular lenses. J Cataract Refract Surg 2008; 34:1312–1317
2. Espindle D, Crawford B, Maxwell A, Rajagopalan K, Barnes R, Harris B, Hileman K. Quality-of-life improvements in cataract patients with bilateral blue light-filtering intraocular lenses: clinical trial. J Cataract Refract Surg 2005; 31:1952–1959
3. Mangione CM, Lee PP, Pitts J, Gutierrez P, Berry S, Hays RD. Psychometric properties of the National Eye Institute Visual Function Questionnaire (NEI-VFQ); the. NEI-VFQ Field Test Investigators. Arch Ophthalmol 1998; 116:1496–1504
4. Bühren J, Pesudovs K, Martin T, Strenger A, Yoon G, Kohnen T. Comparison of optical quality metrics to predict subjective quality of vision after laser in situ keratomileusis. J Cataract Refract Surg 2009; 35:846–855
5. Lundström M, Pesudovs K. Catquest-9SF patient outcomes questionnaire; nine-item short-form Rasch-scaled revision of the Catquest questionnaire. J Cataract Refract Surg 2009; 35:504–513
6. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Using Rasch analysis to revisit the validity of the Cataract TyPE Spec instrument for measuring cataract surgery outcomes. J Cataract Refract Surg 2009; 35:1509–1517
7. Amesbury EC, Grossberg AL, Hong DM, Miller KM. Functional visual outcomes of cataract surgery in patients with 20/20 or better preoperative visual acuity. J Cataract Refract Surg 2009; 35:1505–1508

8. Csaky KG, Richman EA, Ferris FL III. Report from the NEI/FDA Ophthalmic Clinical Trial Design and Endpoints Symposium. Invest Ophthalmol Vis Sci 2008; 49:479–489. Available at: http://www.iovs.org/cgi/reprint/49/2/479. Accessed January 9, 2010
9. Pham HH, Schrag D, O'Malley AS, Wu B, Bach PB. Care patterns in Medicare and their implications for pay for performance. N Engl J Med 2007; 356:1130–1139. Available at: http://content.nejm.org/cgi/reprint/356/11/1130.pdf. Accessed January 9, 2010. Supplementary appendix available at: http://content.nejm.org/cgi/data/356/11/1130/DC1/1. Accessed January 9, 2010
10. Sharma S, Brown GC, Brown MM, Hollands H, Shah GK. The cost-effectiveness of photodynamic therapy for fellow eyes with subfoveal choroidal neovascularization secondary to age-related macular degeneration. Ophthalmology 2001; 108:2051–2059
11. de Boer MR, Moll AC, de Vet HCW, Terwee CB, Völker-Dieben HJM, Van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: a systematic review. Ophthalmic Physiol Opt 2004; 24:257–273
12. Mallinson T. Why measurement matters for measuring patient vision outcomes. Optom Vis Sci 2007; 84:E675–E682. Available at: http://www.optvissci.com/pt/re/ovs/pdfhandler.00006324-200708000-00006.pdf;jsessionid=KZcW21QvZ9vvSbXT2Y5S7dNchsSTlwjP2TbTJFyJJYG2tgLLc2LR!-701738752!181195629!8091!-1. Accessed January 9, 2010
13. Mangione CM, Lee PP, Gutierrez PR, Spritzer K, Berry S, Hays RD. Development of the 25-item National Eye Institute Visual Function Questionnaire; for the National Eye Institute Visual Function Questionnaire Field Test Investigators. Arch Ophthalmol 2001; 119:1050–1058
14. Bach JR, Tilton MC. Life satisfaction and well-being measures in ventilator assisted individuals with traumatic tetraplegia. Arch Phys Med Rehabil 1994; 75:626–632
15. La Grow SJ. Predicting perceived quality of life scores from the National Eye Institute 25-Item Visual Function Questionnaire. Optom Vis Sci 2007; 84:785–788
16. Massof RW. The measurement of vision disability. Optom Vis Sci 2002; 79:516–552. Available at: http://www.optvissci.com/pt/re/ovs/pdfhandler.00006324-200208000-00015.pdf;jsessionid=KZYSBB5kbwjgJ6YYXzpgLywwBK16QcpSb8q3bQGC31s8vY3cdGzj!-701738752!181195629!8091!-1. Accessed January 9, 2010
17. Massof RW. Likert and Guttman scaling of visual function rating scale questionnaires. Ophthalmic Epidemiol 2004; 11:381–399
18. Massof RW. Application of stochastic measurement models to visual function rating scale questionnaires. Ophthalmic Epidemiol 2005; 12:103–124
19. Massof RW. An interval-scaled scoring algorithm for visual function questionnaires. Optom Vis Sci 2007; 84:689–704
20. Massof RW, Ahmadian L. What do different visual function questionnaires measure? Ophthalmic Epidemiol 2007; 14:198–204
21. Wright BD, Masters GN. Rating Scale Analysis (Rasch Measurement Book). Chicago, IL, MESA Press, 1982
22. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, Denmark, Institute of Educational Research, 1960
23. Wright BD, Mok M. Rasch models overview. J Appl Meas 2000; 1:83–106
24. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. Optom Vis Sci 2007; 84:663–674. Available at: http://www.optvissci.com/pt/re/ovs/pdfhandler.00006324-200708000-00005.pdf;jsessionid=KZcJrBxMKHzzcx4VGhCDgGwV2vDSD5HknjcG729LKxDIX1XPvfm1!-701738752!181195629!8091!-1. Accessed January 9, 2010
25. Lamoureux EL, Ferraro JG, Pallant JE, Pesudovs K, Rees G, Keeffe JE. Are standard instruments valid for the assessment of quality of life and symptoms in glaucoma? Optom Vis Sci 2007; 84:789–796
26. Pesudovs K. Patient-centered measurement in ophthalmology–a paradigm shift [commentary]. BMC Ophthalmol 2006; 6:25. Available at: http://www.biomedcentral.com/content/pdf/1471-2415-6-25.pdf. Accessed January 9, 2010
27. Pesudovs K, Garamendi E, Elliott DB. The Quality of Life Impact of Refractive Correction (QIRC) Questionnaire: development and validation. Optom Vis Sci 2004; 81:769–777. Available at: http://www.optvissci.com/pt/re/ovs/pdfhandler.00006324-200410000-00009.pdf;jsessionid=KZlQrwyhmlzpQ0JgPykLn0wJq9FlTyYqrNBBQCG2kTPPcB41qvh1!-1775402713!181195628!8091!-1. Accessed January 9, 2010
28. Pesudovs K, Garamendi E, Elliott DB. The Contact Lens Impact on Quality of Life (CLIQ) Questionnaire: development and validation. Invest Ophthalmol Vis Sci 2006; 47:2789–2796. Available at: http://www.iovs.org/cgi/reprint/47/7/2789. Accessed January 7, 2010
29. Spaeth G, Walt J, Keener J. Evaluation of quality of life for patients with glaucoma. Am J Ophthalmol 2006; 141(suppl):S3–S14
30. Turano KA, Massof RW, Quigley HA. A self-assessment instrument designed for measuring independent mobility in RP patients: generalizability to glaucoma patients. Invest Ophthalmol Vis Sci 2002; 43:2874–2881. Available at: http://www.iovs.org/cgi/reprint/43/9/2874. Accessed January 7, 2010
31. Velozo CA, Lai J-S, Mallinson T, Hauselman E. Maintaining instrument quality while reducing items: application of Rasch analysis to a self-report of visual function. J Outcome Meas 2000-2001; 4:667–680
32. Lamoureux EL, Pallant JF, Pesudovs K, Hassell JB, Keeffe JE. The Impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. Invest Ophthalmol Vis Sci 2006; 47:4732–4741. Available at: http://www.iovs.org/cgi/reprint/47/11/4732. Accessed January 7, 2010
33. Lamoureux EL, Pesudovs K, Pallant JF, Rees G, Hassell JB, Caudle LE, Keeffe JE. An evaluation of the 10-item vision core measure 1 (VCM1) scale (the Core Module of the Vision-Related Quality of Life scale) using Rasch analysis. Ophthalmic Epidemiol 2008; 15:224–233
34. Pesudovs K, Garamendi E, Keeves JP, Elliott DB. The Activities of Daily Vision Scale for cataract surgery outcomes: re-evaluating validity with Rasch analysis. Invest Ophthalmol Vis Sci 2003; 44:2892–2899. Available at: http://www.iovs.org/cgi/reprint/44/7/2892. Accessed January 7, 2010
35. Garamendi E, Pesudovs K, Stevens MJ, Elliott DB. The Refractive Status and Vision Profile: evaluation of psychometric properties and comparison of Rasch and summated Likert-scaling. Vision Res 2006; 46:1375–1383
36. Massof RW, Fletcher DC. Evaluation of the NEI visual functioning questionnaire as an interval measure of visual ability in low vision. Vision Res 2001; 41:397–413
37. Ryan B, Court H, Margrain TH. Measuring low vision service outcomes: Rasch analysis of the seven-item National Eye Institute Visual Function Questionnaire. Optom Vis Sci 2008; 85:112–121
38. Wang C- W. Chan CL-w, Jin H-Y. Psychometric properties of the Chinese version of the 25-item National Eye Institute Visual Function Questionnaire. Optom Vis Sci 2008; 85:1091–1099
39. Labiris G, Katsanos A, Fanariotis M, Tsirouki T, Pefkianaki M, Chatzoulis D, Tsironi E. Psychometric properties of the Greek version of the NEI-VFQ 25. BMC Ophthalmol 2008; 8:4.

Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2287166/pdf/1471-2415-8-4.pdf. Accessed January 7, 2010
40. Owsley C, McGwin G Jr, Jackson GR, Kallies K, Clark M. Cone- and rod-mediated dark adaptation impairment in age-related maculopathy. Ophthalmology 2007; 114:1728–1735
41. Mangione CM, Berry S, Spritzer K, Janz NK, Klein R, Owsley C, Lee PP. Identifying the content area for the 51-item National Eye Institute Visual Function Questionnaire; results from focus groups with visually impaired persons. Arch Ophthalmol 1998; 116:227–233
42. Kirkwood BJ, Pesudovs K, Latimer P, Coster DJ. The efficacy of a nurse-led preoperative cataract assessment and postoperative care clinic. Med J Aust 2006; 184:278–281
43. Rubin GS, Bandeen-Roche K, Huang G-H, Muñoz B, Schein OD, Fried LP, West SK. The association of multiple visual impairments with self-reported visual disability: SEE project; for the SEE Project Team. Invest Ophthalmol Vis Sci 2001; 42:64–72. Available at: http://www.iovs.org/cgi/reprint/42/1/64. Accessed January 7, 2010
44. Andrich D. A rating scale formulation for ordered response categories. Psychometrika 1978; 43:561–573
45. Massof RW, Rubin GS. Visual function assessment questionnaires. Surv Ophthalmol 2001; 45:531–548
46. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas 2002; 3:85–106
47. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. Arch Phys Med Rehabil 1989; 70:857–860
48. Duncan PW, Bode RK, Lai SM, Perera S. Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale; Glycine Antagonist in Neuroprotection Americas Investigators. Arch Phys Med Rehabil 2003; 84:950–963
49. Velozo CA, Kielhofner G, Lai JS. The use of Rasch analysis to produce scale-free measurement of functional ability. Am J Occup Ther 1999; 53:83–90
50. Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. London, UK, Lawrence Erlbaum Associates, 2001
51. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3:205–231
52. Wright BD, Linacre JM. Reasonable mean-square fit values. Rasch Meas Trans 1994; 8:370. Available at: http://www.rasch.org/rmt/rmt83b.htm. Accessed January 7, 2010
53. Smith RM, Suh KK. Rasch fit statistics as a test of the invariance of item parameter estimates. J Appl Meas 2003; 4:153–163
54. Wang W-C, Yao G, Tsai Y-J, Wang J-D, Hsieh C- L. Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. Qual Life Res 2006; 15:607–620
55. Ünver YB, Yavuz GA, Sinclair SH. Interactive, computer-based, self-reported, visual function questionnaire: the PalmPilot-VFQ. Eye 2009; 23:1572–1581
56. Swamy BN, Chia E-M, Wang JJ, Rochtchina E, Mitchell P. Correlation between vision- and health-related quality of life scores. Acta Ophthalmol (Oxf) 2009; 87:335–339
57. Revicki D, Rentz A, Harnam N, Thomas V, Lanzetta P. Reliability and validity of the National Eye Institute Visual Function Questionnaire-25 in patients with age-related macular degeneration. In press, Invest Ophthalmol Vis Sci 2010
58. Hariprasad SM, Mieler WF, Grassi M, Green JL, Jager RD, Miller L. Vision-related quality of life in patients with diabetic macular oedema. Br J Ophthalmol 2008; 92:89–92
59. Lamoureux EL, Hooper CY, Lim L, Pallant JF, Hunt N, Keeffe JE, Guymer RH. Impact of cataract surgery on quality of life in patients with early age-related macular degeneration. Optom Vis Sci 2007; 84:683–688
60. Pesudovs K, Coster DJ. Cataract surgery reduces subjective visual disability. Aust N Z J Ophthalmol 1997; 25(4):S3–S5
61. Pesudovs K, Coster DJ. An instrument for assessment of subjective visual disability in cataract patients. Br J Ophthalmol 1998; 82:617–624. Available at: http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1722619&blobtype=pdf. Accessed January 7, 2010
62. Castro-Costa E, Dewey M, Stewart R, Banerjee S, Huppert F, Carlos Mendonca-Lima C, Bula C, Reisches F, Wancata J, Ritchie K, Tsolaki M, Mateos R, Prince M. Ascertaining late-life depressive symptoms in Europe: an evaluation of the survey version of the EURO-D scale in 10 nations; the SHARE project. Int J Methods Psychiatr Res 2008; 17:12–29
63. van Nispen RMA, Knol DL, Neve HJ, van Rens GHMB. A multilevel item response theory model was investigated for longitudinal vision-related quality-of-life data. In press, J Clin Epidemiol 2010
64. Stelmack JA, Massof RW. Using the VA LV VFQ-48 and LV VFQ-20 in low vision rehabilitation. Optom Vis Sci 2007; 84:705–709
65. Johnson ME, Murphy PJ. Measurement of ocular surface irritation on a linear interval scale with the ocular comfort index. Invest Ophthalmol Vis Sci 2007; 48:4451–4458. Available at: http://www.iovs.org/cgi/reprint/48/10/4451. Accessed January 7, 2010

First Author:
Konrad Pesudovs, PhD

NH&MRC Centre for Clinical Eye Research, Department of Optometry and Vision Science, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, South Australia, Australia