Clinical and Epidemiologic Research

The Psychometric Validity of the NEI VFQ-25 for Use in a Low-Vision Population

Manjula Marella,[1] *Konrad Pesudovs,*[2] *Jill E. Keeffe,*[1] *Patricia M. O'Connor,*[1] *Gwyneth Rees,*[1] *and Ecosse L. Lamoureux*[1]

Purpose. To determine the psychometric validity of the National Eye Institute-Visual Function Questionnaire (NEI VFQ-25) and its subscale structure for use in people with low vision.

Methods. Two hundred thirty-two participants completed the NEI VFQ-25. Rasch analysis was used to test the psychometric performance of the questionnaire and each subscale. Factor models were hypothesized and tested with confirmatory factor analysis (CFA) and subsequently validated with Rasch analysis.

Results. For the overall scale, two rating scales had to be dichotomized and three misfitting items removed to improve fit to the Rasch model. There was evidence of multidimensionality, indicating that the scale would benefit from scale splitting. For the NEI VFQ-25 subscale structure, six of the original 12 subscales could not fit the Rasch model because of item insufficiency (fewer than two items) and the remaining six displayed poor item fit characteristics indicating that the NEI VFQ-25 does not have a viable subscale structure. CFA supported a two-factor model with visual functioning (10 items) and socioemotional (9 items) scales. Most goodness-of-fit statistics were within the recommended range of values. The factor loadings of items on their respective scales were statistically significant ($P < 0.001$) and ranged between 0.59 and 0.84. The two scales individually fitted the Rasch model and were found to be unidimensional with adequate psychometric characteristics.

Conclusions. The native NEI VFQ-25 is a better performing instrument when split into visual functioning and socioemotional scales. These scales possess valid parameters for assessment of the impact of low vision in this population. (*Invest Ophthalmol Vis Sci.* 2010;51:2878–2884) DOI:10.1167/iovs.09-4494

The 25-item National Eye Institute Visual Function Questionnaire (NEI VFQ-25) is one of the most widely used of the visual function questionnaires.[1] Reduced from the original 51-item version,[2,3] the NEI VFQ-25 has reliability and validity comparable to that of the longer version (as does the 39-item version [NEI VFQ-39] which represents the NEI VFQ-25 plus an appendix of extra items).[1] It has been used in large population–based eye surveys[4–6] and has been validated in several languages.[7–9]

The NEI VFQ-25 was designed to reduce the respondent's burden (compared with the NEI VFQ-39) and increase its suitability and validity for clinical trials.[1] Although the validity of the NEI VFQ-25 is beyond question in a conventional sense, there are newer psychometric methods that give greater insight into questionnaire validity. Rasch analysis, a form of Item Response Theory (IRT), has been used by several investigators to validate and re-engineer questionnaires and their subscales.[10] Questionnaires such as the Activities of Daily Living Scale (ADLS),[11] the Visual Function questionnaires (VF-14[12] and VF-11[13]), Catquest,[14] and the Impact of Vision Impairment (IVI)[15] have been revalidated. Massof and Fletcher[16] applied Rasch analysis to 27 functioning items of the original NEI VFQ in a low-vision population and found multidimensionality (i.e., more than one construct being measured; a fundamental flaw, as a single score should represent a single construct). The NEI VFQ-39 was shown to be sensitive to the impact of low-vision rehabilitation by Stelmack et al.,[17] using Rasch analysis. However, the most widely used version, the NEI VFQ-25, has not been tested by Rasch analysis for validity in the low-vision population. Furthermore, the validity of its subscales is unknown, as they are yet to be psychometrically assessed by Rasch analysis. This lack of subscale validity is a concern, as many of the NEI VFQ-25 subscales contain only a few items, and other questionnaire subscales with few items have been shown to be invalid.[14,18]

Using Rasch analysis, we assessed whether the total and subscale scores of the NEI VFQ-25 are psychometrically valid to be used in a low-vision population. We hypothesized that the Rasch analysis would show that overall and subscale constructs of the NEI VFQ-25 are not psychometrically optimal for use in a low-vision population. For the overall trait, we will consider remedial actions such as category collapsing, item deletion, and splitting the scale to achieve fit. For the original subscale structure, if invalid, we will propose different factor models after the removal of misfitting items, and these will be based on the hypothesized constructs associated with the remaining items. New structure models will then be assessed by using confirmatory factor analysis (CFA) and Rasch analysis.

Methods

Participants

Participants who had low vision diagnosed at tertiary public eye clinics at the Royal Victorian Eye and Ear Hospital (RVEEH; Melbourne) were recruited for the study between 2001 and 2002. Eligibility criteria included an ability to converse in English, visual acuity <6/12, and age 18 years or older. Individuals who agreed to participate signed a written consent form. The study was approved by the RVEEH Human Research and Ethics Committee and adhered to the tenets of the Declaration of Helsinki.

Sociodemographic, vision, general health, and NEI VFQ-25 data for each participant were obtained with interviewer-administered question-

From the [1]Centre for Eye Research Australia, the Royal Victorian Eye and Ear Hospital, University of Melbourne, Melbourne, VIC, Australia; and [2]National Health and Medical Research Council (NH&MRC) Centre for Clinical Eye Research-Flinders University and Flinders Medical Centre, Adelaide, SA, Australia.

Submitted for publication August 17, 2009; revised December 22, 2009, and January 11, 2010; accepted January 11, 2010.

Disclosure: **M. Marella**, None; **K. Pesudovs**, None; **J.E. Keeffe**, None; **P.M. O'Connor**, None; **G. Rees**, None; **E.L. Lamoureux**, None

Corresponding author: Ecosse L. Lamoureux, Centre for Eye Research Australia, Department of Ophthalmology, University of Melbourne, Locked Bag 8, East Melbourne VIC, 8002, Australia; ecosse@unimelb.edu.au.

TABLE 1. Items of the NEI VFQ-25, Subscales and Their Responses

No.	Items	Subscales	Response	Missing Data (%)	Floor Effect (%)*
1	General health	General health	Quality	0	12.1
2	General vision	General vision	Quality	2.6	31.5
3	Worry about eyesight	Mental health	Frequency	0	26.3
4	Pain around eyes	Ocular pain	Quality	0	3.9
5	Reading normal newsprint	Near vision	Difficulty	0.4	53.9
6	Seeing well up close	Near vision	Difficulty	1.7	25.9
7	Finding objects on crowded shelf	Near vision	Difficulty	0.4	17.2
8	Street signs	Distance vision	Difficulty	2.6	26.3
9	Going downstairs at night	Distance vision	Difficulty	2.2	7.3
10	Seeing objects off to side	Peripheral vision	Difficulty	1.3	7.3
11	Seeing how people react	Social function	Difficulty	0.9	16.8
12	Matching clothes	Color vision	Difficulty	1.3	4.7
13	Visiting others	Social function	Difficulty	7.8	13.4
14	Going out to movies/plays	Distance vision	Difficulty	21.1	35.3
15	Driving in daylight	Driving	Difficulty	82.2	2.6
16	Driving in difficult conditions	Driving	Difficulty	86.2	6.5
17	Accomplish less	Role limitations	Frequency	0	13.8
18	Limited endurance	Role limitations	Frequency	0.4	33.6
19	Amount of time in pain	Ocular pain	Frequency	0	65.5
20	Stay home most of the time	Dependency	Agreement	0.9	39.7
21	Frustrated	Mental health	Agreement	2.2	20.7
22	No control	Mental health	Agreement	2.6	21.1
23	Rely too much on others' words	Dependency	Agreement	3.9	34.5
24	Need much help from others	Dependency	Agreement	1.3	38.4
25	Embarrassment	Mental health	Agreement	2.2	52.6

* Floor effect is percentage of answers in the no difficulty/none of the time/definitely false.

naires. After an eye examination, the cause of vision loss, presenting and best-corrected visual acuity (VA), and visual field results were recorded.

The NEI VFQ-25 Questionnaire

Table 1 lists the items and subscales of the NEI VFQ-25. Twelve items relate to the difficulty of the activities, with five response options ranging from *no difficulty at all*, to *stopped doing because of eyesight.* A sixth option *stopped doing for other reasons* or *not interested in doing* was considered missing data and was not included in the analyses. Four items relating to frequency of the problems due to vision loss are rated on a five-category scale of options ranging from *all the time* to *none of the time*. Six items are statements of agreement to problems associated with vision loss on a five-category scale ranging from *definitely true* to *definitely false*. The response options of the 12 items requiring difficulty ratings and the item *worry about eyesight* were reversed for the Rasch analysis, so that the measures would have the same polarity (i.e., better and worse were consistent across all items).

In addition to an overall score, items are grouped into 12 subscales. These represent general health (one item), general vision (one item), ocular pain (two items), near vision (three items), distance vision (three items), driving (two items), peripheral vision (one item), color vision (one item), role limitations (two items), dependency (three items), social function (two items), and mental health (four items).

Statistical Analysis

Demographic and clinical data of the participants were analyzed (SPSS ver. 17 for Windows; SPSS Inc., Chicago, IL). Rasch analysis[19,20] was performed (Winsteps, ver. 3.67, Chicago, IL)[21] to determine the validity and reliability of the NEI VFQ-25. As there are six different types of rating scales, we used four Andrich rating-scale models[19] (one for each item group) to obtain the estimates of the required visual ability of each item, perceived visual ability of each person, and the category thresholds for each response categories. For each of the item types with different rating scales, we first investigated evidence of disordered thresholds, as this indicates whether the participants could not reliably discriminate between the response categories. Combining adjacent categories is often the solution to disordered thresholds.[22]

Once response category performance was satisfactory, person and item measures were examined for fit to the Rasch model, with an unconditional maximum-likelihood estimation routine. Rasch analysis locates item difficulty and person ability on a logit scale (log of odds). How well the data fit the model was evaluated by the item fit statistics infit and outfit. The information-weighted (infit) statistic is more sensitive to the pattern of responses to person-targeted items and less sensitive to the presence of outliers; therefore, it is the main fit statistic reported herein. The outlier-sensitive (outfit) statistic is sensitive to unexpected behavior by persons or items far from the subject's ability level. In the mean square (MNSQ) form, fit statistics show variance in the data with an expected value of 1.0. MNSQ values less than 1.0 indicate that the items are too predictable, thereby suggesting redundancy. Values of more than 1.0 suggest unpredictability due to noise in the data and are considered to be misfitting. Values between 0.7 and 1.3 are considered acceptable.[23,24] These values represent 30% less or more variance than expected for the item.

The person separation index is the ratio of the variance in the person measures for the sample to the average error in estimating these measures. It is a measure of how broadly the persons could be distinguished into statistically distinct levels. The person separation reliability coefficient describes the reliability of the scale to discriminate between the persons of different abilities. A person separation index of $\geq$2.0 or a reliability value of $\geq$0.8 represents the minimum acceptable level of separation.[20,23] A value of 0.8 is equivalent to a person separation ratio (G) of 2, which means that there are three strata [strata = $(4G + 1)/3$], or significantly different levels, of person ability that can be distinguished by the items.[20,25] Targeting is a method of assessing how well the difficulty of the items in the scale suits the ability of the sample. Suitability can be assessed by inspecting the person-item maps or numerically using the mean scores for person and item measures. Effective targeting is evident when the person and item means are close to each other.[23,26]

To test the hypothesis that the NEI VFQ-25 measures a single underlying construct (unidimensionality) we conducted a principal components analysis (PCA) of the residuals (difference between the observed and expected responses).[27,28] Data are considered unidimensional if most of the variance is explained by the principal component and there is no significant explanation of the residual variance by the

Table 2. Characteristics of the 232 Study Participants

	n	%
Age, y		
≤50	34	14.7
>50	198	85.3
Sex		
Male	122	52.6
Female	110	47.4
Presenting visual acuity (better eye)		
<6/12–6/18 (mild vision impairment)	84	36.2
<6/18–6/60 (moderate vision impairment)	97	41.8
<6/60 (severe vision impairment)	51	22.0
Main cause of vision loss		
Retinal diseases	103	44.4
Glaucoma	38	16.4
Cataract	32	13.8
Other	59	25.4

contrasts to the principal component. The variance explained by the principal component for the empiric calculation should be comparable to that of the model and should be >60%.[28] Furthermore, the unexplained variance by the contrasts should be <2.0 Eigenvalue units, which is close to that seen with random data.

For further validation, we tested differential item functioning (DIF), which assesses whether the items have different meanings for the different groups of the sample. DIF is tested for a range of variables: age, sex, ocular comorbidity, and level of vision loss. The raw differences in item calibration between groups were examined to identify DIF. DIF was considered absent if it was less than 0.50 logits, minimal but probably inconsequential if it ranged between 0.50 and 1.0 logits, and notable if it was >1.0 logit.[29,30]

The 12 subscales of the NEI VFQ-25 were then analyzed separately by using the same procedures and criteria as that used to analyze the overall scale. However, six subscales did not fit the Rasch model due to item insufficiency. Hence, new factor models were hypothesized based on common underlying themes that classify most of the remaining items. These new models were then assessed by using the CFA and Rasch analysis.

Using the Rasch calibrated person measures, hypothesized factor models were evaluated by CFA (performed with AMOS, ver 16; SPSS Science, Chicago, IL). This analysis allows the assessment of the overall model fit, testing the relationship between the observed variables and their underlying latent constructs in the model. To determine the adequacy of model fit with the data, we used the following fit indices: (1) χ^2, (2) the goodness-of-fit index (GFI), (3) the adjusted goodness-of-fit index (AGFI), (4) the comparative fit index (CFI), (5) the Tucker-Lewis index (TLI), and (6) the root mean square error of approximation (RMSEA). A nonsignificant χ^2 probability value indicates a good model fit. However, χ^2 is sensitive to sample size. To address this concern, a relative χ^2 is used (ratio of χ^2 to degrees of freedom, χ^2/df) with a recommended range of 1.0 to 2.0.[31] For GFI, AGFI, CFI, and TLI values, <0.90 indicates lack of fit, between 0.90 and 0.95 indicates reasonable fit, and between 0.95 and 1.00 indicates good fit.[32–34] The RMSEA values must be ≤0.05 to indicate good fit. Values between 0.05 and 0.08 indicate reasonable fit.[33,34] The scale structure (with the best fit characteristics) was then examined for validity and unidimensionality with Rasch analysis (Winsteps, ver. 3.67).[21]

Results

Two hundred thirty-two participants (52.6% men) with a mean age 67.3 years (SD ±16.3) were recruited. Most (78%) had mild or moderate vision loss (presenting visual acuity in the better eye <6/12–6/60). The most common cause of vision loss was retinal disease (44.4%). The participants' sociodemographic characteristics are summarized in Table 2.

Rasch Analysis

Overall Score. Total scores (calculated as recommended by the developers) for 25 items ranging from 11.49 to 62.50 with a mean score (±SD) of 42.14 ± 8.70. The item information curves for six ratings scales are given in Figure 1. The initial fit of the NEI VFQ-25 to the Rasch model showed disordered thresholds for two rating scales. The rating scale with response options ranging from *all the time* to *none of the time* had underutilization of the middle categories. This underutilization was overcome by collapsing the categories to the dichotomous scale 1 (always) and 2 (never). The rating scale with response options ranged from *definitely true* to *definitely false*, the category *not sure* was not used at all by our participants, and there was an overlap of thresholds between categories 3 and 4. Therefore, not sure was coded as a missing category and the responses were recorded as 1 (true) and 2 (false).

After category collapsing, three items were found to misfit (general health, pain around eyes, driving in difficult conditions) with infit mean scores >1.3. These three items were removed iteratively in the following order: *pain around eyes*, *general health*, and *driving in difficult conditions*. Removal of these items improved the fit of the scale to the Rasch model. Fit statistics of the remaining items are given in Table 3.

The mean (±SD) of the person measures was 0.18 ± 1.49 logits. The separation index for person measures was 3.28, with reliability of 0.91. Using these separation indices with the formula of Wright and Masters,[20] our sample could be divided into five statistically distinct strata. The person-item map (Fig. 2) displays the person and item measures in logits and indicates good targeting. The top half of the graph is the distribution of persons and the lower half is the distribution of the item measures.

The PCA of item residuals revealed that the variance explained by measures for the empiric calculation (62.1%) was comparable to that explained by the model (63.9%). The unexplained variance by the first contrast accounted for 2.9 Eigenvalue units, suggesting presence of a second dimension in the scale. No further contrasts exceeded 2.0 Eigenvalue units. Eight items loaded (correlation, >0.4) positively onto the first contrast and belonged to dependency (three items), mental health (three items), and role limitations (two items) subscales. This suggests that these eight items cannot be grouped with other items in the scale to measure a single latent trait.

There was no DIF found for any of the items in the study indicating that the items were interpreted similarly across subgroups of the sample. The criterion validity of the NEI VFQ-25 was tested by assessing its ability to discriminate between participants of different levels of visual impairment: mild, moderate, and severe (as defined in Table 2). There was a significant difference between the three groups (ANOVA, $F = 44.37$; $P < 0.0001$), with poorer visual acuity associated with poorer quality of life (1.10, −0.02, and −0.97).

Overall, the results from the Rasch analysis showed that NEI VFQ-25 met the criteria for psychometric validation after the categories were collapsed and the three items removed. However, there was evidence of some multidimensionality with the PCA of the residuals.

Subscale Structure. Analysis of the native subscale structure showed that 6 of the 12 subscales of the NEI VFQ-25 did not fit the Rasch model because of item insufficiency (<2 items). The remaining six subscales displayed poor item fit characteristics, indicating that the NEI VFQ-25 does not have a viable subscale structure. Collectively, these findings confirm our initial hypothesis that the NEI VFQ-25 subscale structure is performing suboptimally.

We hypothesized two models from the remaining 22 items—namely, a two-factor model and a three-factor model. It was evident from PCA that the dependency, mental health, and

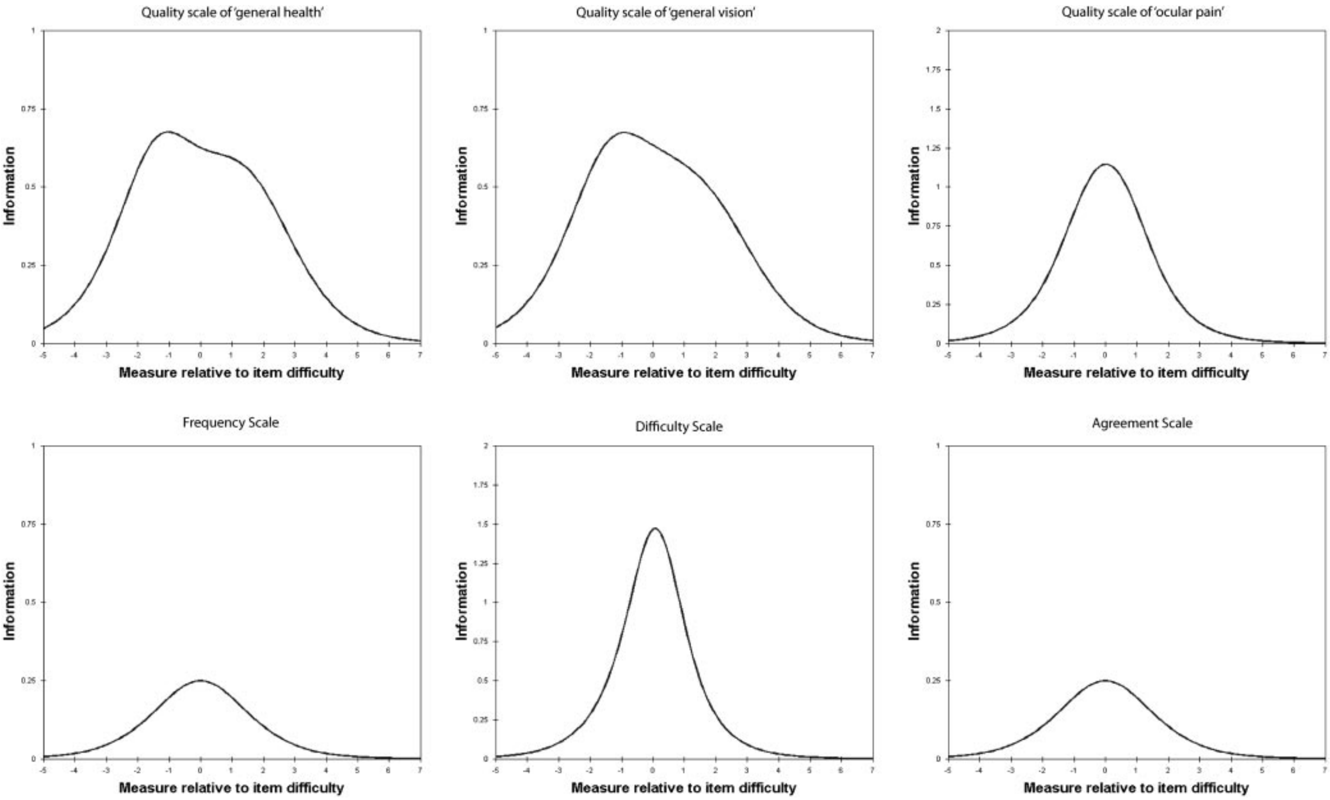

FIGURE 1. Item information curves for six response scales.

role limitations subscales form a second dimension. Therefore, we combined those nine items into a socioemotional scale. The remaining subscales near vision, distance vision, peripheral vision, color vision, and social functioning were combined into a visual functioning scale. These two formed a two-factor model. We also hypothesized a three-factor model. Since mobility was shown to be different subscale from reading in other scales such as the impact of vision impairment (IVI)[35] and the activity breakdown structure (ABS),[36] we further classified the visual functioning scale into *reading and accessing information* (five items), and *mobility and social functioning* (five items). These two scales along with the socioemotional items formed a three-factor model. The items *general vision* and *pain* could not be fitted with any of the proposed scales and were excluded from the analysis. For the item *driving in daylight*, 83.2% of the participants chose the option, stopped doing this for other reasons or not interested in doing this, and these responses were con-

TABLE 3. Fit Statistics after Removal of Misfitting Items

Items	Measure	Error	Infit MNSQ	Outfit MNSQ
General vision	1.98	0.10	1.13	1.07
Reading normal newsprint	1.32	0.08	1.27	1.05
No control	1.22	0.17	0.79	0.66
Frustrated	0.99	0.17	0.87	0.81
Going out to movies/plays	0.54	0.09	1.23	1.12
Accomplish less	0.39	0.16	0.74	0.61
Rely too much on others' words	0.33	0.16	0.75	0.63
Street signs	0.33	0.08	1.21	1.30
Going downstairs at night	0.17	0.08	1.25	1.70
Need much help from others	0.13	0.16	0.85	0.76
Seeing well up close	0.13	0.07	1.18	1.08
Stay home most of the time	0.10	0.16	0.87	0.79
Worry about eyesight	−0.02	0.15	1.08	1.34
Finding objects on crowded shelf	−0.03	0.07	0.87	0.85
Seeing how people react	−0.29	0.07	1.03	0.92
Seeing objects off to side	−0.40	0.07	1.13	1.07
Visiting others	−0.52	0.08	1.33	1.12
Limited in endurance	−0.55	0.16	0.81	0.67
Embarrassment	−0.61	0.16	0.94	0.92
Driving in daylight	−1.38	0.28	1.32	1.05
Matching clothes	−1.43	0.09	1.12	0.84
Amount of time in pain	−2.42	0.22	1.05	0.91

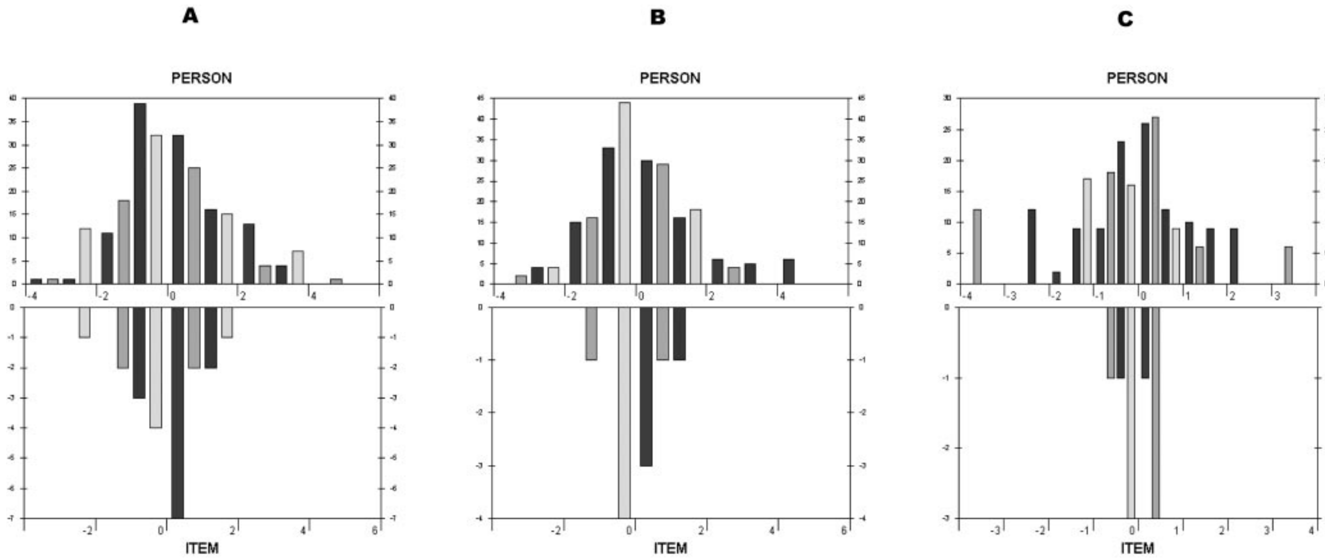

Figure 2. The targeting maps for (**A**) the overall NEI VFQ-25; (**B**) *vision functioning* scale; and (**C**) *socioemotional* scale showing the distribution of the person and item measures.

sidered missing data. Hence, this item could not be included in the new structure assessment.

Confirmatory Factor Analysis

Before CFA was conducted, the assumptions of—a continuous and normally distributed sample, no systematic missing data and a sufficiently large sample size—were tested.[37] The Rasch-converted measures were found to be continuous and normally distributed (Kolmogorov-Smirnov [K-S] test, $P > 0.05$). Participants were removed in the event of any missing data. As there are no specific criteria for adequate sample size, we decided that a ratio of at least five participants to one item was needed. The sample of 135 participants for the remaining 19 items was adequate for CFA analyses.

The selected goodness-of-fit statistics for the two proposed models are shown in Table 4. The fit indices, χ^2/df, CFI, TLI, and RMSEA were identical and within the recommended range for both models, indicating good fit of the NEI VFQ-25 data for the proposed models. The GFI (0.86) and AGFI (0.82) were slightly less than the recommended value 0.90 for both models. However, there was a high correlation (0.98) between the scales *reading and accessing information* and *mobility and social* in the three-factor model. High correlation indicated a strong overlap between the latent traits of these two scales. Of the two proposed factor models, the two-factor model with the latent traits of visual functioning (10 items) and socioemotional status (9 items) displayed better fit statistics. The standardized regression weights for the 19 items were all statistically significant ($P < 0.001$) and ranged between 0.67 and 0.84 for visual functioning and 0.59 and 0.81 for socioemotional status. The interfactor correlation between the two scales was 0.87, indicating an overlap between the factors.

To further test the strong correlation between the two models we formally tested a one-factor model (with 19 items) to see whether it would measure a single trait: quality of life. Fit statistics of the one-factor model showed poor fit to the data (Table 4). Although the χ^2 statistic was not significant, the other fit statistics, GFI (0.72), AGFI (0.64), CFI (0.86), TLI (0.84), and RMSEA (0.115), were worse than the recommended values, indicating misfit. Overall, the best fitting model for these data was marginally the two-factor model, with visual functioning and socioemotional scales consisting of 19 items.

Fit of the Scales to the Rasch Model

The psychometric validity of the two-factor model identified by the CFA was subsequently assessed by using Rasch analysis. Both scales showed satisfactory fit to the Rasch model with acceptable separation indices, reliabilities, and targeting parameters (Table 5, Figs. 2B, 2C). There were no misfitting items in both scales. The PCA of the residuals for the two scales revealed that they are unidimensional constructs for visual functioning and socioemo-

Table 4. Comparative Goodness-of-Fit Statistics for Two Hypothesised Models of the Revised NEI VFQ-25.

		Values		
Fit Indices	**Recommended Value**	**Two-Factor Model**	**Three-Factor Model**	**One-Factor Model**
χ^2	N/A	237.810	236.613	399.249
df	N/A	147	145	151
χ^2/df	≤2.00	1.618	1.632	2.644
GFI	≥0.90	0.86	0.86	0.72
AGF	≥0.90	0.82	0.82	0.64
CFI	≥0.90	0.95	0.95	0.86
TLI	≥0.90	0.94	0.94	0.84
RMSEA		0.067	0.068	0.115
(90% confidence intervals)	≤0.08	(0.052–0.081)	(0.052–0.083)	(0.101–0.129)

Table 5. Rasch Analysis Fit Statistics of the Two-Factor Model Scales

Scales	Items in Scale (n)	Misfitting Items (n)	Person Separation Index	Person Separation Reliability	Mean ± SD Person Measure (logits)	Principal Component Analysis (Eigenvalue)
Visual functioning	10	None	2.59	0.87	0.19 ± 1.44	1.7
Socioemotional	9	None	2.00	0.80	−0.16 ± 2.12	1.8

tional status, respectively. All the items were free of DIF. There was a significant difference between the three visual impairment groups (ANOVA, $F = 47.34$ and 46.64, respectively, for visual functioning and socioemotional scales; $P < 0.0001$) with poorer visual acuity associated with poor quality of life for both scales: 1.10, −0.01, and −0.94 for visual functioning and 1.89, 0.59, and −1.17 for socioemotional scales.

Discussion

The purpose of this study was to determine the validity of the NEI VFQ-25 for use as an overall or subscale score of quality of life in a population with low vision. As an overall scale, NEI VFQ-25 had satisfactory psychometric properties according to a Rasch analysis. However, as anticipated, the subscale structure in its original form was psychometrically suboptimal, which warranted re-engineering according to CFA and Rasch analysis. Because of the high interfactor correlation found in the three-factor model, the two-factor model appeared to be reliable to assess visual functioning and socioemotional traits in this population. Our findings suggest that the original 12-subscale structure of the NEI VFQ-25 has limited psychometric validity and could be replaced by our proposed two-factor model to provide a better understanding of the impact of vision loss on the person.

Items related to *general health*, *pain around the eyes*, and *driving* did not fit the overall scale. The same items were found to be misfitting in a similar study by this group[38] in a cataract population. The item pain around the eyes displayed high skewness and deviation from the expected model. In our population, more than 80% did not drive, which resulted in high levels of missing data for both the driving items. This result suggests that driving may not be a valid or relevant question for low-vision persons, who by definition are not legally allowed to drive. Our findings with driving items are similar to those of Broman et al.[39] and Globe et al.[7] Massof and Fletcher[16] have also reported in their study using the original version of the 51-item NEI VFQ that very few low-vision participants in their sample drove. This item should thus be removed from the NEI VFQ when administered to a population with low vision. In addition, three items: general vision, pain, and driving in difficult conditions could not be grouped with any of the hypothesized models for the CFA, thus further reducing the NEI VFQ to two scales consisting of 10 and 9 items in visual functioning and socioemotional scales, respectively. The item reduction of the scale is one of the strengths of the study as the two mini scales now further reduce the administration time.

In this study, the categories for two rating scales (10 items) had to be collapsed to a dichotomous scale. Dichotomizing the rating scale is a concern, as it may fail to discriminate between the spectrum of disability compared to a three- or four-rating scale and may add skewness to the data.[40] On the other hand, our participants could effectively discriminate between the five-category rating scale for items relating to the difficulty of the activities due to vision loss. This finding contrasts with previous studies in which categories had to be collapsed to a three- or four-category response scale.[11,15,41,42] Discrepancies between studies could be due to the specificity of eye conditions, severity of visual impairment, or variability in the protocol of questionnaire administration.

Valid scales are important in the study of the impact of low vision and measure the outcome of interventions. Of the two proposed structures, the two-factor model displayed better psychometric properties. However, the socioemotional scale had borderline precision with regard to person separation. Person separation depends on the range of abilities in the sample and the length of the rating scale.[28] Since the rating scale for socioemotional scale had been collapsed to two categories, the person separation values weakened.

We found that the mobility and reading items grouped to form a single scale. This finding is in contrast with other scales such as the IVI[35] and the ABS,[36] where mobility has been found to be a separate subscale. Discrepancies between studies could be related to differences in vision characteristics of the study samples or the specific nature of the items. For example, in the IVI study,[35] almost two thirds of the sample had severe visual impairment (visual acuity <6/60). In our study, more than three fourths of the sample had moderate vision loss (presenting visual acuity ranging between 6/18 and 6/60), which suggests that mobility may not have caused the same level of difficulty, or there may not be sufficient mobility items in the NEI-VFQ to form a valid measure of mobility. On the other hand, socioemotional status, although with borderline precision, was found to be an independent scale. This is an important finding, as it allows assessment of the socioemotional distress of people with low vision. When measured using this scale, lower or negative person measures indicate higher emotional distress.

There was a strong overlap ($r = 0.87$) between the two scales in this study indicating that the two variables could be sharing an inherent trait, such as quality of life. However, they do not fit as a single-factor model. This overlap could possibly be because the emotional status depends on the visual functioning and could be affected when the latter is impaired.

The newly proposed scales demonstrated acceptable reliability, validity, and unidimensionality, indicating that they measure underlying traits of visual functioning and socioemotional status. Our findings are indirectly supported by previous work with the original 51-item NEI VFQ.[3,16] Using classic test theory, Mangione et al.[3] found strong interscale correlations (0.75–0.85) for near vision, distance vision, and social function subscales (which constitute our visual functioning scale), indicating that these items share common properties. In that same study, role limitations, dependency, and mental health subscales (grouped under our socioemotional scale) also showed good interscale correlation scores (0.72–0.77). In addition, Massof and Fletcher[16] found the same groups of scales contributing to latent traits when analyzed independently with Rasch analysis.

Our study had two limitations. The sample consisted of mostly people with retinal diseases (44.4%). Future research should assess the psychometric validity of the NEI VFQ-25 in a sample with more varied range of diseases. The other limitation was that, despite an adequate sample for CFA, a larger sample size would have given more strength to our findings. We recommend testing the validity of the NEI VFQ-25 on the respective sample it is being used for.

Other investigators who wish to use the 19-item NEI VFQ can use our validation data to convert raw scores into Rasch scores without having to perform Rasch analysis. This conversion mainly holds for patients with complete data. Rather than an overall score, we suggest splitting the NEI-VFQ into two mini scales. Our spreadsheet has been prepared to convert raw scores to Rasch scores for both the scales.

In conclusion, our findings indicate that the original NEI VFQ-25 is not psychometrically optimal for assessing either overall vision-related functioning or subtraits in a low-vision population. Rather, we found that items grouped under visual functioning and socioemotional traits were psychometrically valid constructs. Future studies are needed to substantiate our findings and evaluate the sensitivity of these new scales for rehabilitation programs or eye care interventions for people with low vision.

References

1. Mangione CM, Lee PP, Gutierrez PR, Spritzer K, Berry S, Hays RD. Development of the 25-item National Eye Institute Visual Function Questionnaire. *Arch Ophthalmol.* 2001;119:1050–1058.
2. Mangione CM, Berry S, Spritzer K, et al. 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.
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). NEI-VFQ Field Test Investigators. *Arch Ophthalmol.* 1998;116:1496–1504.
4. Clemons TE, Chew EY, Bressler SB, McBee W. National Eye Institute Visual Function Questionnaire in the Age-Related Eye Disease Study (AREDS): AREDS Report No. 10. *Arch Ophthalmol.* 2003;121:211–217.
5. Cole SR, Beck RW, Moke PS, Gal RL, Long DT. The National Eye Institute Visual Function Questionnaire: experience of the ONTT; Optic Neuritis Treatment Trial. *Invest Ophthalmol Vis Sci.* 2000; 41:1017–1021.
6. Klein R, Moss SE, Klein BE, Gutierrez P, Mangione CM. The NEI-VFQ-25 in people with long-term type 1 diabetes mellitus: the Wisconsin Epidemiologic Study of Diabetic Retinopathy. *Arch Ophthalmol.* 2001;119:733–740.
7. Globe D, Varma R, Azen SP, Paz S, Yu E, Preston-Martin S. Psychometric performance of the NEI VFQ-25 in visually normal Latinos: the Los Angeles Latino Eye Study. *Invest Ophthalmol Vis Sci.* 2003;44:1470–1478.
8. Toker E, Onal S, Eraslan M, Eyriparmak M. The Turkish version of the National Eye Institute Refractive Error Quality of Life Instrument: translation, validity and reliability. *Qual Life Res.* 2008;17:1269–1276.
9. Wang CW, Chan CL, Jin HY. Psychometric properties of the Chinese version of the 25-item National Eye Institute Visual Function Questionnaire. *Optom Vis Sci.* 2008;85:1091–1099.
10. Pesudovs K. Patient-centered measurement in ophthalmology: a paradigm shift. *BMC Ophthalmol.* 2006;6:25.
11. 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.
12. Velozo CA, Lai JS, 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;4:667–680.
13. Lamoureux EL, Pesudovs K, Thumboo J, Saw SM, Wong TY. An evaluation of the reliability and validity of the visual functioning questionnaire (VF-11) using Rasch analysis in an Asian population. *Invest Ophthalmol Vis Sci.* 2009;50:2607–2613.
14. Lundstrom 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.
15. 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.
16. 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.
17. Stelmack JA, Stelmack TR, Massof RW. Measuring low-vision rehabilitation outcomes with the NEI VFQ-25. *Invest Ophthalmol Vis Sci.* 2002;43:2859–2868.
18. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Rasch Analysis of the Quality of Life and Vision Function Questionnaire. *Optom Vis Sci.* Published online September 7, 2009.
19. Andrich D. A Rating formulation for ordered response categories. *Psychometrika.* 1978;43:561–573.
20. Wright B, Masters G. *Rating Scale Analysis: Rasch Measurement.* Chicago: MESA Press; 1982.
21. Linacre JM. WINSTEPS Rasch measurement [computer program]. Chicago: Winstepscom; 2005.
22. Andrich D, Wright B. Rasch sensitivity and Thurstone insensitivity to graded responses. *Rasch Measure Trans.* 1994;8:382.
23. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci.* 2007; 84:663–674.
24. Wright BD. Reasonable mean-square fit values. In: Wright BD, Linacre JM, eds. *Rasch Measurement Transactions.* Chicago: MESA Press; 1996:360.
25. Smith EV Jr. Evidence for the reliability of measures and validity of measure interpretation: a Rasch measurement perspective. *J Appl Meas.* 2001;2:281–311.
26. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Med Care.* 2004;42:I17–24.
27. 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.
28. Linacre JM, Wright BD. *A User's Guide to Winsteps Rasch-Model Computer Program.* Chicago: MESA Press: 2001.
29. Wright BD, Douglas GA. Best test design and selftailored testing. Research Memorandum No. 19. Chicago: Statistical Laboratory, Department of Education, University of Chicago; 1975.
30. Wright BD, Douglas GA. Best test design and selftailored testing. Research Memorandum No. 21. Chicago: Statistical Laboratory, Department of Education, University of Chicago; 1976.
31. Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. *Multivariate Data Analysis.* 6th ed. Englewoods Cliffs, NJ: Prentice Hall; 2005.
32. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull.* 1990;107:238–246.
33. Byrne B. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming.* Mahwah, NJ: Lawrence Erlbaum Associates; 2001.
34. Kline R. *Principles and Practices of Structural Equation Modeling.* 2nd ed. New York: Guilford Press; 2005.
35. Lamoureux EL, Pallant JF, Pesudovs K, Rees G, Hassell JB, Keeffe JE. The impact of vision impairment questionnaire: an assessment of its domain structure using confirmatory factor analysis and Rasch analysis. *Invest Ophthalmol Vis Sci.* 2007;48:1001–1006.
36. Massof RW, Hsu CT, Baker FH, et al. Visual disability variables. II: The difficulty of tasks for a sample of low-vision patients. *Arch Phys Med Rehabil.* 2005;86:954–967.
37. Kaplan D. *Structural Equation Modeling: Foundations and Extensions.* 2nd ed. Los Angeles: SAGE; 2009:xi, 255 p.
38. Pesudovs K, Gothwal VK, Wright TA, Lamoureux EL. National Eye Institute-Visual Function Questionnaire: what does it measure? *J Cataract Refract Surg.* In press.
39. Broman AT, Munoz B, West SK, et al. Psychometric properties of the 25-item NEI-VFQ in a Hispanic population: Proyecto VER. *Invest Ophthalmol Vis Sci.* 2001;42:606–613.
40. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* Mahwah, NJ: L Erlbaum; 2001:xxiv, 255 p.
41. Stelmack JA, Szlyk JP, Stelmack TR, et al. Psychometric properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire. *Invest Ophthalmol Vis Sci.* 2004;45:3919–3928.
42. Turano KA, Geruschat DR, Stahl JW, Massof RW. Perceived visual ability for independent mobility in persons with retinitis pigmentosa. *Invest Ophthalmol Vis Sci.* 1999;40:865–877.