

Psychometric Properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire

Joan A. Stelmack,^{1,2,3} Janet P. Szlyk,^{3,4} Thomas R. Stelmack,^{2,3,4} Paulette Demers-Turco,^{5,6} R. Tracy Williams,^{2,7,8} D'Anna Moran,¹ and Robert W. Massof⁹

PURPOSE. To describe psychometric properties of a self-report questionnaire, the Veterans Affairs (VA) Low-Vision Visual Functioning Questionnaire (LV VFQ-48), which was designed to measure the difficulty visually impaired persons have performing daily activities and to evaluate low-vision outcomes.

METHODS. The VA LV VFQ-48 was administered by telephone interview to subjects with visual acuity ranging from near normal to total blindness at five sites in the VA and private sector. Rasch analysis with the Andrich rating scale model was applied to difficulty ratings from 367 subjects, to evaluate measurement properties of the instrument.

RESULTS. High intercenter correlations for item measure estimates (intraclass correlation coefficient [ICC] = 0.97) justified pooling the data from these sites. The person measure fit statistics (mean square residuals) confirm that the data fit the assumptions of the model. The item measure fit statistics indicate that responses to 19% of the items were confounded by factors other than visual ability. The separation reliabilities for pooled data (0.94 for persons and 0.98 for items) demonstrate that the estimated measures discriminate persons and items well along the visual ability dimension. ICCs for test-retest data (0.98 for items and 0.84 for persons) confirm temporal stability. Subjects used the rating categories in the same way at all five centers. Ratings of slight and moderate difficulty were used interchangeably, suggesting that the instrument could be modified to a 4-point scale including not difficult, slightly/moderately difficult, extremely difficult, and impossible. Fifty additional subjects were administered the questionnaire with a

4-point scale to confirm that the scale was used in the same way when there were four rather than five difficulty ratings.

CONCLUSIONS. The VA LV VFQ-48 is valid and reliable and has the range and precision necessary to measure visual ability of low-vision patients with moderate to severe vision loss across diverse clinical settings. (*Invest Ophthalmol Vis Sci.* 2004;45:3919-3928) DOI:10.1167/iovs.04-0208

Low vision is defined as any chronic visual impairment that interferes with everyday function and is not correctable by ordinary spectacles or contact lenses.¹ The loss of visual abilities can have a profound effect on the life of an individual, by limiting performance of even simple everyday tasks, such as dressing, eating, writing, traveling from place to place, and communicating with others.² Persons with low vision may feel frustrated and restricted in their lifestyles when they are unable to perform these everyday activities independently. Low-vision services enhance the use of remaining vision, increase independence, and improve the quality of life of persons with chronic visual impairments.³ Low-vision service delivery begins with an extensive interview to learn about the patient's problems, needs, and goals. To promote cost-effective low-vision service delivery and to ensure that all relevant issues are covered, a standard assessment tool is needed to capture the patient's self-report of difficulty performing activities at the initial interview, and the change in difficulty performing these activities in the community after rehabilitation. In this article, we describe the psychometric properties of a new self-report visual functioning questionnaire that was designed to serve both as a patient evaluation tool and as a rehabilitation outcome measure.

Visual functioning questionnaires include a set of questions that assess the performance of daily living activities.⁴ The activities being rated are referred to as items. Items work in synchrony to create a scale. The number of response choices included on visual functioning questionnaires ranges from as few as 2 to as many as 10. Responses are usually ordered by rank. Items are usually assigned a numerical score corresponding to the patient's rating. Most visual functioning questionnaires employ Likert scoring, which is simply a linear transformation of the average response rank across items. Although it is hoped that Likert scores are monotonic with the person trait the instrument is attempting to measure, there are no means to confirm that they are without a normative measurement model. Therefore, we used the Rasch analysis to estimate interval visual ability scales from Likert-type visual function rating scales to test construct validity.⁵ We have used Rasch analysis to compare the difficulty that patients in two VA low-vision programs experienced in performing activities before and after rehabilitation.⁶ Other investigators have used Rasch analysis to validate a questionnaire to assess mobility in patients with visual field loss,^{7,8} an activities of daily living (ADL) index (the Melbourne Low Vision ADL Index),⁹ and a questionnaire to measure functional vision performance of visually impaired children (LV Prasad-Functional Vision Questionnaire).¹⁰ Most recently, Rasch analysis has been used to revalidate or improve on existing instruments, including the

From the ¹Blind Rehabilitation Center, Edward Hines VA Hospital, Hines, Illinois; the ²Illinois College of Optometry, Chicago, Illinois; the ³Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago College of Medicine, Chicago, Illinois; the ⁴Jessie Brown VA Medical Center, Chicago, Illinois; the ⁵Massachusetts Eye and Ear Infirmary, Boston, Massachusetts; the ⁶Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts; the ⁷Lois and Edwin Deicke Center for Vision Rehabilitation, Wheaton, Illinois; the ⁸Department of Ophthalmology, Loyola University School of Medicine, Maywood, Illinois; and the ⁹Wilmer Eye Institute, Johns Hopkins University, Baltimore, Maryland.

Supported by the Office of Research and Development, Rehabilitation Research and Development Service of the Department of Veterans Affairs, Illinois Society for Prevention of Blindness, Massachusetts Eye and Ear Infirmary Joel A. Kraut Teaching and Research Fund (PD-T), Max Goldenberg Foundation (RTW), and Grant EY-012045 of the National Eye Institute (RWM).

Submitted for publication February 25, 2004; revised June 21, 2004; accepted June 28, 2004.

Disclosure: **J.A. Stelmack**, None; **J.P. Szlyk**, None; **T.R. Stelmack**, None; **P. Demers-Turco**, None; **R.T. Williams**, None; **D. Moran**, None; **R.W. Massof**, None

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Joan A. Stelmack, Blind Rehabilitation Center (124), Edward Hines Hospital, Hines, IL 60141; joan.stelmack@med.va.gov.

National Eye Institute Visual Functioning Questionnaire (NEI VFQ),¹¹ the Visual Function self-report (VF-14),^{12,13} and the Activities of Daily Vision Scale.¹⁴

Although there are currently more than a dozen visual functioning questionnaires, most of the instruments that were available 5 years ago, when we began our outcomes studies, were developed for clinical trials of eye disease treatments that restore or improve vision (such as the removal of cataracts).^{3,4} None of these instruments is used routinely in delivery of low-vision services. Different visual functioning questionnaires were developed with different applications in mind. For example, the NEI VFQ-25 was developed with the purpose of measuring vision-specific quality of life in clinical research,^{15,16} whereas the VF-14 was developed with the purpose of measuring functional limitations caused by cataract and the outcomes of cataract surgery.¹⁷ Although many items from visual functioning questionnaires, including the Activities of Daily Vision Scale, VF-14, and NEI VFQ-25, could be used to measure low-vision rehabilitation outcomes, the number of items on each instrument is not sufficient to evaluate the range of activities that are a problem for persons with vision loss.¹⁸ Based on our review of these questionnaires and their development, we determined that an additional instrument was needed to measure the outcomes of low-vision service delivery.

METHODS

The development of the 48-item Veterans Affairs (VA) Low-Vision Visual Functioning Questionnaire (LV VFQ-48) using a modified Delphi method and preliminary analysis of data (before full sample) with the Andrich Rating scale model is described in two previous publications.^{19,20} Clinicians, rehabilitation specialists, scientists, research staff and persons with vision loss contributed to item development and selection. Service providers were asked to draw on their many years of experience in patient care to identify patient's needs, activities that are commonly addressed in each rehabilitation program, and expected outcomes. The preferred practice patterns and clinical guidelines for VA blind rehabilitation services, optometry, ophthalmology, and occupational therapy were examined. An extensive literature review was conducted. Development, validation, and sensitivity to change of the existing visual functioning questionnaires and quality of life instruments were reviewed.^{19,20} In addition to generation of new items, questions and items from existing instruments were incorporated in the VA LV VFQ-48 or modified to improve their usefulness. Additional information was available from a previously conducted study that included structured interviews with persons with low vision to determine patients' needs for low-vision devices.²¹ The 48 items chosen by the research team for the initial version of the VA LV VFQ-48 are listed in Table 1.^{19,20}

The initial version of the VA LV VFQ includes four questions. Question 1 is asked about all 48 items: "Is it difficult to . . . ?" Response choices include: *not difficult*, *slightly difficult*, *moderately difficult*, *extremely difficult*, *impossible*, and *do not do it for nonvisual reasons (which is scored as missing data)*. Question 2, "Is it (difficult) because of your vision?", is asked if the subject responds that it is difficult to perform an activity. Response choices are *yes* or *no*. If it is difficult to perform the activity because of vision loss, question 3 is asked, "Do you want training?" Responses are also *yes* or *no*. Patients who are able to perform the activities are asked in question 4 how they perform the activity. Question 4 is "How do you usually . . . ?" Response choices include: *own eyes or eye glasses*, *vision devices/techniques* (e.g., magnifier) *other senses/nonvisual devices* (e.g., cane), *someone helps me*, and *not applicable*.

Administration of the VA LV VFQ-48

The VA LV VFQ-48 was administered by telephone before rehabilitation. Two interviewers located at the Department of Ophthalmology

TABLE 1. Items Included in the 48-Item Field Test Version VA LV VFQ

1. Physically get dressed
2. Keep clean
3. Identify medicine
4. Tell time
5. Identify money
6. Match clothes
7. Groom yourself
8. Identify food on a plate
9. Eat and drink neatly
10. Fix a snack
11. Prepare meals
12. Use appliance dials
13. Clean the house
14. Handle finances
15. Make out a check
16. Take a message
17. Find something on a crowded shelf
18. Find public restrooms
19. Get around indoors in places you know
20. Get around outdoors in places you know
21. Get around in unfamiliar places
22. Go out At night
23. Go down steps in dim light
24. Adjust to bright light
25. Cross streets at traffic lights
26. Use public transportation
27. Get around in a crowd
28. Avoid bumping into things
29. Recognize persons up close
30. Recognize persons from across the room
31. Read street signs and store names
32. Read headlines
33. Read menus
34. Read newspaper or magazine articles
35. Read mail
36. Read small print on package labels
37. Read print on TV
38. Keep your place while reading
39. Watch TV
40. Play table and card games
41. See photos
42. Work on your favorite hobby
43. Go to movies
44. Go to spectator events
45. Play sports
46. Do yard work
47. Sign your name
48. Read signs

(University of Illinois at Chicago) conducted interviews of patients from the Visual Impairment Center to Optimize Remaining Sight (VICTORS). Three interviewers located at the Hines Blind Rehabilitation Center conducted interviews for the other four sites. Administration time for all four questions on the VA LV VFQ-48 was 25 to 35 minutes.

Subjects

English-speaking patients were recruited from five clinical sites: the Hines Blind Rehabilitation Center (Hines BRC), Hines VA Outpatient Low Vision Rehabilitation/Eye Clinic (Hines LV Clinic), Jessie Brown VA Medical Center VICTORS Program (VICTORS), Lois and Edwin Deicke Center for Visual Rehabilitation (Deicke Center) and the Vision Rehabilitation Center at Massachusetts Eye and Ear Infirmary (MEEI). These sites were chosen to include a representative sample of patients with low vision who are typically referred or self-refer to low-vision clinics. The VA sites include a suburban hospital (Hines) and an inner-city medical center (Jessie Brown VA Medical Center). Included are regional VA programs targeting services to legally blind veterans (Hines BRC) and visually impaired veterans (VICTORS), as well as a

TABLE 2. Description of Subjects

Site Name	Hines BRC	Hines LV Clinic	VICTORS	Deicke	MEEI	Total	Verification
Total subjects (<i>n</i>)	95/367 (26)	66/367 (18)	50/367 (13)	98/367 (27)	58/367 (16)	367 (100)	50
Males (<i>n</i>)	90/95 (95)	59/66 (89)	49/50 (98)	44/98 (45)	28/58 (48)	270/367 (74)	44/50 (88)
Mean age (y)	69 ± 11.6	71 ± 11.5	70 ± 9.1	77 ± 11.4	70 ± 17.0	71 ± 12.0	75 ± 10.0
Range	(42-87)	(39-88)	(51-90)	(37-95)	(27-91)	(27-95)	(50-90)

Data are the number of persons in the total group, with the percentage of the total in parentheses.

local VA hospital outpatient low vision/eye clinic (Hines LV Clinic). The private sector sites chosen include a not-for-profit agency located in the Chicago suburbs (Deicke Center) and a hospital-based rehabilitation service located in a major city (MEEI), the latter of which is affiliated with both a department of ophthalmology and a department of physical medicine and rehabilitation.

Clinical staff at each program site excluded patients who would not be able to participate in the study. Reasons for exclusion included active major depression, cognitive loss, terminal illness, or other serious health conditions. The study was conducted in compliance with the tenets of the Declaration of Helsinki for research in human subjects.²² Informed consent was received based on the requirements of the local institutional review board (IRB) protocols approved at each participating site with oversight from the IRBs at Hines Hospital and the University of Illinois at Chicago.

The recruitment goal was a convenience sample of at least 50 subjects per site. Recruitment goals were exceeded at four of the centers, so the sample size for the initial analysis was 367 patients. A description of the subjects from each site is provided in Table 2. Distribution of subjects by visual acuity is presented in Table 3. Primary diagnoses of subjects are listed by site in Table 4.

The test was administered a second time 3 to 4 weeks later to 30 subjects from the Hines BRC and the Hines LV Clinic. These subjects included 28 men and 2 women (mean age, 73 ± 9.4 (SD) years; range, 49-88 years). Primary diagnosis of these subjects was macular degeneration (50%), diabetic retinopathy (27%), glaucoma (13%), neurologic disorders (7%), and cataracts (3%). Habitual visual acuity was near-

normal impairment (13%), moderate impairment (13%), severe impairment (54%), profound impairment (13%), and nearly or absolutely no light perception (7%).

As will be described in the Results section, subject ratings of slightly and moderately difficult were used interchangeably so that the categories could be collapsed to a 4-point scale (not difficult, slightly/moderately difficult, extremely difficult, and impossible), with no loss of information. Therefore, the questionnaire was administered to an additional 50 subjects recruited from the Hines BRC, Hines LV Clinic, and Deicke Center offering only four difficulty rating categories rather than five, to determine the validity of collapsing response categories in the data analysis. The 50 subjects in the validation study are described separately (as Verification) in Tables 2, 3, and 4.

Statistical Analysis

Rasch analysis^{23,24} was performed using Winsteps²⁵ on difficulty ratings for the combined data from all five sites ($N = 367$). Rasch analysis also was performed separately on the data sets from each of the five sites, on the set of retest data from the Hines BRC and Hines LV Clinic, and on the additional 50 subjects from the Hines BRC, Hines LV Clinic, and Deicke Center who responded to the version of the instrument having four difficulty rating categories. Rasch analysis, using the Andrich Rating scale model,²⁴ provides estimates of the visual ability of each person (α_n) the required visual ability (i.e., inherent difficulty) of each item (ρ_i) and the threshold (τ_c) for each response category (i.e., $\alpha - \rho$, where response probabilities for neighboring response categories are equal), on an interval logit scale.

TABLE 3. Habitual Visual Acuity of Subjects

Site Name	Hines BRC	Hines LV Clinic	VICTORS	Deicke	MEEI	Total	Verification
Normal VA							
(VA > 20/40)*	0/95	6/66	9/50	1/98	6/58	22/367	3/50
LogMAR < 0.301 (%)†	0	9	18	1	10	6	6
SD LogMAR	—	0.070	0.068	—	0.044	0.072	0.108
Near-normal impairment							
(20/40-20/60)*	6/95	8/66	20/50	32/98	11/58	77/367	5/50
0.301-0.477 (%)†	6	12	40	33	19	21	10
SD LogMAR	0.070	0.065	0.077	0.065	0.056	0.073	0.089
Moderate impairment							
(20/70-20/160)*	21/95	14/66	9/50	35/98	26/58	105/367	11/50
0.544-0.903 (%)†	22	21	18	36	45	28	22
SD LogMAR	0.138	0.096	0.072	0.118	0.168	0.137	0.121
Severe impairment							
(20/200-20/400)*	57/95	28/66	12/50	25/98	15/58	137/367	28/50
1.000-1.301 (%)†	61	42	24	25	26	38	56
SD LogMAR	0.128	0.130	0.136	0.122	0.116	0.127	0.142
Profound impairment							
(20/500-20/1000)	7/95	6/66	0/50	4/98	0/58	17/367	2/50
1.398-1.698 (%)†	7	9	0	4	0	5	4
SD LogMAR	0.053	0.074	—	0.063	—	0.346	0.156
Near-total impairment/total impairment (NLP)							
(VA < 20/1000)*	4/95	4/66	0/50	1/98	0/58	9/367	1/50
LogMAR > 1.698 (%)†	4	7	0	1	0	2	2

Based on evaluation of best eye by World Health Organization classification. NLP, no light perception.

* Snellen VA.

† LogMAR VA.

TABLE 4. Primary Eye Diagnosis of Subjects

Site Name	Hines BRC	Hines LV Clinic	VICTORS	Deicke	MEEI	Total	Verification
Macular degenerations	58/95 (61)	19/66 (29)	25/50 (50)	73/98 (74)	32/58 (55)	207/367 (57)	28/50 (56)
Other retinal degenerations	3/95 (3)	2/66 (3)	1/50 (2)	2/98 (2)	5/58 (9)	13/367 (4)	2/50 (4)
Diabetic retinopathy	12/95 (13)	22/66 (33)	10/50 (20)	9/98 (10)	5/58 (9)	58/367 (15)	8/50 (16)
Retinal detachment	1/95 (1)	0 (0)	0 (0)	2/98 (2)	0 (0)	3/367 (1)	1/50 (2)
Other retinal vascular disorders	2/95 (2)	0 (0)	0 (0)	0 (0)	2/58 (3)	4/367 (2)	1/50 (2)
Strabismus and other disorders of binocular eye movements	1/95 (1)	0 (0)	0 (0)	0 (0)	1/58 (2)	2/367 (0)	0 (0)
Glaucoma	11/95 (12)	15/66 (23)	4/50 (8)	7/98 (7)	5/58 (9)	42/367 (11)	3/50 (6)
Neurological disorders	5/95 (5)	7/66 (11)	10/50 (20)	5/98 (5)	6/58 (10)	33/367 (9)	0 (0)
Cataracts	1/95 (1)	1/66 (1)	0 (0)	0 (0)	2/58 (3)	4/367 (1)	6/50 (12)
Trauma	1/95 (1)	0 (0)	0 (0)	0 (0)	0 (0)	1/367 (0)	1/50 (2)

Data are the number of persons in the total group, with the percentage of the total in parentheses.

The Andrich rating scale model is a normative probabilistic conjoint measurement model.²⁴ It is normative, because it conforms to the structural requirements of axiomatic measurement theory.²⁶ If the VA LV VFQ is valid as a measurement instrument, the data must agree with the expectations of the model. If the VA LV VFQ is to be useful as a measurement instrument, the estimated measures must be precise and reliable. To evaluate measurement precision, the analysis includes standard errors of each estimate and measurement reliability coefficients. To test measurement validity, the analysis includes normative measurement model fit statistics and estimates of conjoint structural coherence. To evaluate test-retest reliability and to evaluate the validity of collapsing response categories, we used the intraclass correlation coefficient (ICC).²⁷

Estimates of Response Probabilities

The Andrich rating scale model is

$$\pi_{nix} = \frac{\exp[x(\alpha_n - \rho_i) - \sum_{j=1}^x \tau_j]}{1 + \sum_{k=1}^m \exp[x(\alpha_n - \rho_i) - \sum_{j=1}^k \tau_j]}$$

where π_{nix} is the probability of person n responding with difficulty rating x to item i for $m + 1$ response categories ($x = 0$ to m).^{23,24}

RESULTS

The results reported in this article are limited to the difficulty ratings of question 1 obtained before rehabilitation.

Figure 1 illustrates estimates of π for each difficulty-rating category as a function of functional reserve, $\alpha - \rho$ (person measure - item measure), based on the response matrix for all 367 subjects from the five sites combined.

Note that there is no value of $\alpha - \rho$ for which difficulty rating 2 is the most probable response. This observation suggests that patients used response category 2 idiosyncratically, which would imply that patients cannot reliably discriminate more than four categories of difficulty. This interpretation of the probability curves is echoed in Guttman's coefficient of reproducibility,²⁸ or coherence, for each difficulty-rating category: 83% for category 1, 21% for category 2, 28% for category 3, 38% for category 4, and 76% for category 5 (a coherence of 100% means that there is perfect order in the response matrix, i.e., all items agree in their ordering of person ability and all persons agree in their ordering of item difficulty; a coherence of 0% means that there is no ordered pattern in the response matrix). The Guttman coefficient of reproducibility across all

responses is 57%. Category 2 is the least coherent, which means it is least able to discriminate among persons of different ability and among items of different required ability.

Because of the poor performance of rating categories 2 and 3, we combined those two response categories and reanalyzed the data with four difficulty rating categories instead of five. Figure 2 illustrates response probabilities for the recoded data set as a function of $\alpha - \rho$.

In this case, there is a value of $\alpha - \rho$ for which each of the difficulty rating categories is the most probable response. Guttman's coefficients of reproducibility for the recoded difficulty rating categories are 78% for category 1, 44% for category 2, 38% for category 3, and 73% for category 4. The overall Guttman coefficient of reproducibility for the recoded data set is 60%.

To test the validity of recoding the difficulty rating categories, we obtained data from an additional 50 subjects and used four difficulty response categories instead of five. Figure 3 illustrates the estimated response probability functions for each of the four rating categories as a function of $\alpha - \rho$ (solid curves).

The Guttman coefficients of reproducibility for these data are 80% for category 1, 41% for category 2, 43% for category 3, and 78% for category 4. The overall Guttman coefficient of reproducibility is 62%. The response probability functions from Figure 2 are reproduced in Figure 3 for comparison (broken curves). The agreement between the two sets of probability functions confirms the validity of combining response categories 2 and 3 in the recoded data set.

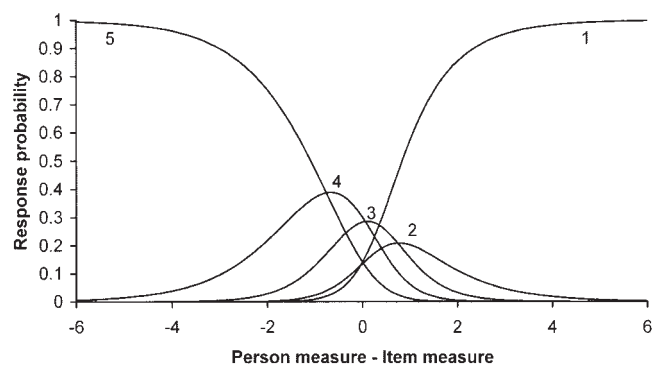


FIGURE 1. Estimates of response probabilities using five response categories. The π for each difficulty rating category is estimated as a function of functional reserve, $\alpha - \rho$ (person measure - item measure), based on the response matrix for all 367 subjects from the five sites combined.

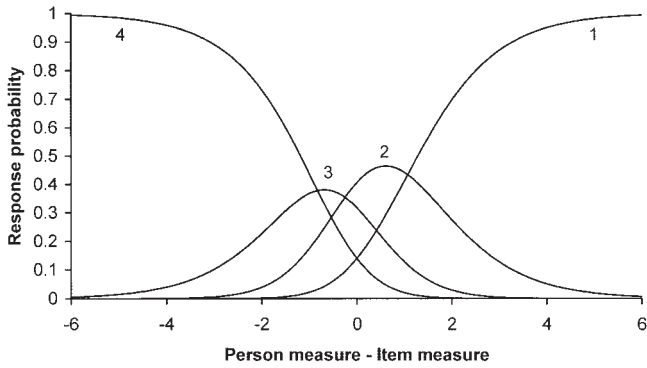


FIGURE 2. Figure 2 illustrates response probabilities for the recoded data set as a function of $\alpha - \rho$ when rating categories 2 and 3 are combined.

Estimates of Person and Item Measures

Winsteps estimates person and item measures, α_n and ρ_i , for each person and item, by an unconditional maximum-likelihood estimation routine.²⁵ Both person and item measures are reported as log odds ratios (logits), which are on an interval scale. Item measure estimations from the separate Rasch analyses of each subset of data agree across the five sites (ICC = 0.97). Therefore, we combined the data of all sites to estimate item measures. The origin of the logit scale is defined to be the average required ability across all items. By definition, the average estimated value of ρ_i for the 48 items in the VA LV VFQ is zero. The SD of ρ_i is 1.06 with a range of -3.07 (least difficult) to 1.72 (most difficult). The root mean square standard error of the estimate over all items is 0.09. The precision of the item measures is represented by a separation reliability coefficient, which is the ratio of the adjusted variance to the observed variance in the item measure distribution. The adjusted variance is the difference between the observed variance and the mean square SE (i.e., adjusted variance is an estimate of the actual variance in ρ across items). Separation reliability, which ranges from 0 to 1, is similar to Cronbach's α and is an index of precision with which differences in ρ between items can be discriminated.¹³ The item measure separation reliability is 0.993.

The estimated values of α range from -1.89 (least able) to 5.21 (most able). The mean (\pm SD) of the person measure distribution is 0.98 ± 1.07 . The root mean square SE of the estimate for all patients is 0.27. The person measure separation reliability is 0.93. Figure 4 compares cumulative distributions of α for each of the five sites.

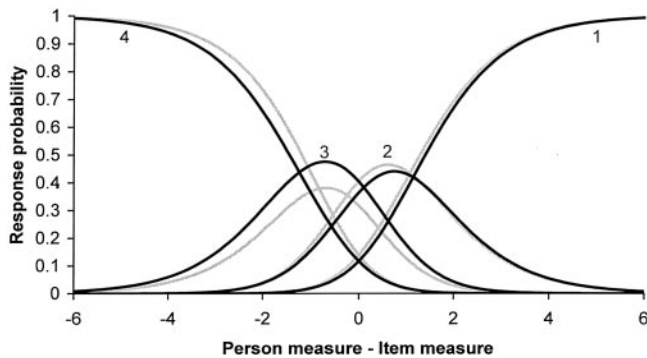


FIGURE 3. Figure 3 illustrates the estimated response probability functions for each of the four rating categories as a function of $\alpha - \rho$ (solid curves) for 50 subjects. For comparison, the probability curves in Figure 2 are presented as gray curves.

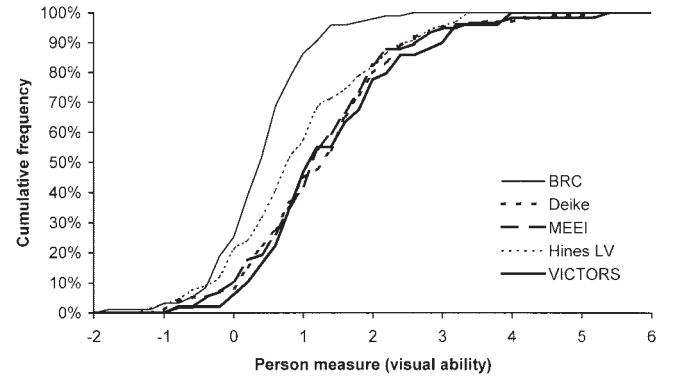


FIGURE 4. Figure 4 compares cumulative distributions of α for each of the five sites, demonstrating that the α distributions for Deicke, MEEI, and VICTORS are the same, whereas the α distribution for the BRC is significantly different from the other distributions.

The α distributions for Deicke Center, MEEI, and VICTORS are the same (Kolmogorov-Smirnov [K-S] test, $P = 0.88, 0.95, 0.99$, respectively).²⁹ The α distribution for the Hines BRC is significantly different from the other distributions (K-S, $P < 0.001$). The distribution for the Hines LV Clinic has a lower mean than the MEEI, Deicke Center, and VICTORS distributions, but the differences do not achieve statistical significance, given the multiple comparisons (K-S, $P = 0.04, 0.167, 0.168$, respectively).

Figure 5 illustrates the person-item map. The vertical axis of the logit scale is visual ability ranging from least able (lowest

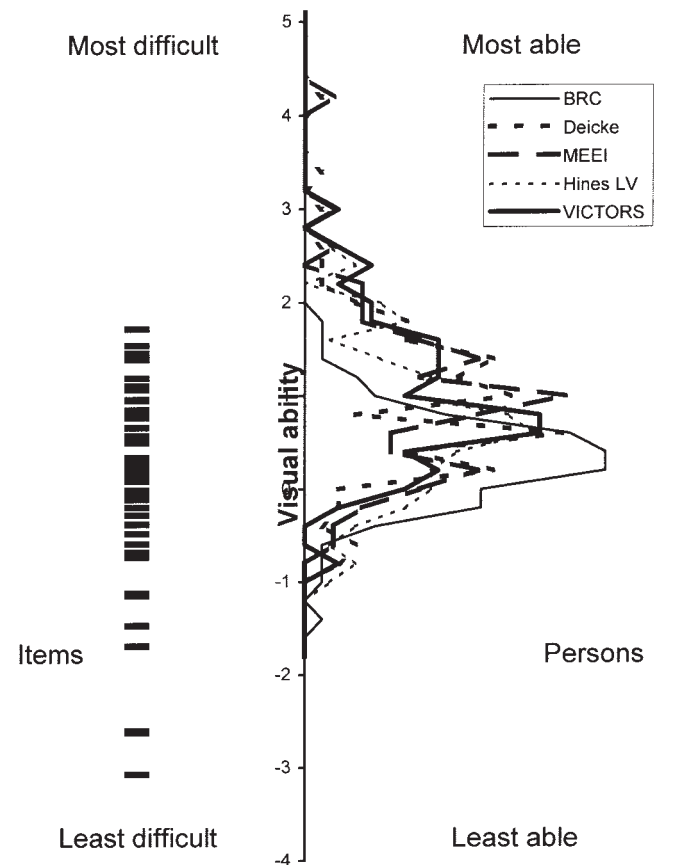


FIGURE 5. The Rasch Person Item Map displays a scale created from measurement of patients' abilities to perform activities of daily living (right) and the visual ability needed to perform each activity (left).

value of α ; at the bottom) to most able (highest value of α ; at the top). The curves on the right are the frequency distributions of person measures for each of the five sites. The horizontal lines to the left of the axis represent the distribution of item measures ranging from least difficult (lowest value of ρ) to most difficult (highest value of ρ). The item measure distribution is well centered on the Hines BRC person measure distribution, but the person measure distributions for the other sites extend to higher visual abilities than are covered by the items (this means that the rating scale is doing most of the work of separating persons having greater ability). The items are not evenly spaced, and many of the items have similar item measures. Therefore, from a measurement perspective, there is high redundancy in the item measures and the instrument performs less well in discriminating among people with high ability than among people with lower ability.

Validity of Person and Item Measure Estimates

The Andrich Rating scale model is a normative measurement model. It is based on a Rasch logistic model and conforms to the tenets of axiomatic measurement theory.^{23,24} Consequently, if the estimated measures from the VA LV VFQ difficulty ratings are valid, the pattern of patient responses to the items must agree with the expectations of the model within the limits of statistical error.³⁰ Goodness of fit is evaluated with weighted mean square residual errors across items for each person, across persons for each item, and across all person-item encounters for the instrument.³⁰ Two different weighting schemes are used. First, for each person-item encounter, the squared residual error is normalized to the expected variance. This normalized squared residual, which is expected to be distributed as χ^2 , is called the outfit statistic, because it is sensitive to outlying errors.³⁰ For the second weighting scheme, the mean square residual is normalized to the average expected variance. This normalized mean squared residual, which also is expected to be distributed as χ^2 , is called the infit statistic, because it is most representative of inlying errors.³⁰ Because both fit statistics are expected to be distributed as χ^2 , the weighted mean squares are transformed to an expected standard normal distribution using a Wilson-Hilferty transformation.³¹ Thus, the transformed outfit and infit values for each item and person are presented as z -scores with an expected mean of 0 and an expected SD of 1.

Table 5 lists the estimated measures (ρ), the SE of the estimate, the infit and outfit mean squares (MNSQ), and the z -score transformations (ZSTD) of the infit and outfit mean squares for each item.

Figure 6 illustrates a scatterplot of the infit versus outfit mean square z -scores for the 48 items.

The box encloses the values that are within ± 2 SD of the expected value (0). Negative z -scores below and to the left of the box indicate that the residual errors for those items are less than would be expected from normally distributed measurement error. Positive z -scores above and to the right of the box indicate that the residual errors for those items are greater than would be expected from normally distributed measurement error. Of the 48 items, 13 (27%) have outfit mean squares, 10 (21%) have infit mean squares, and 9 (19%) items have both outfit and infit mean squares that are greater than the expected value by >2 SD. Two of the items, work on favorite hobby and use public transportation, have outfit and infit mean squares that exceed the expected value by >5 SD. Other items with both outfit and infit mean squares that exceed 2 SD include, identify medicine, handle finances, go out at night, get around in a crowd, avoid bumping into things, read headlines, and sign your name. At the other extreme, 8 (17%) items have outfit mean squares and 10 (21%) items have infit mean squares

that are less than the expected value by >2 SD. For two of the items, get around in unfamiliar places and see photos, both the outfit and infit mean squares are less than the expected value by >4 SD.

Figure 7 illustrates the distribution of z -scores for person measure infit mean squares along the x -axis and the distribution of person measures (α) along the y -axis. The dashed vertical lines enclose the region that is ± 2 SD from the expected value. Eleven percent of the patients have infit mean squares greater than the expected value by ≥ 2 SD (compared with 2.5% for a normal distribution) and 7% of the patients have infit mean squares less than the expected value by ≥ 2 SD (compared with 2.5% for a normal distribution). Two patients from VICTORS and two from Hines LV Clinic are grossly misfitting with infit mean squares greater than the expected value by >4 SD. Their response patterns to the items indicate that some confounding variable(s), and not the variable of interest, is governing the responses of these patients.

The average infit and outfit mean squares across persons are 1.02 and 1.04, respectively, compared with an expected value of 1.00. In terms of z -scores, the average person infit and outfit are both 0.1, compared with an expected value of 0.0, and the standard deviations are 1.7 and 1.6, respectively, compared with an expected value of 1.0. The average infit and outfit mean squares across items are 1.01 and 1.03, respectively, compared with an expected value of 1.00. In terms of z -scores, the average item infit and outfit are 0.1 and 0.3, respectively, compared with an expected value of 0.0, and the standard deviations are 2.8 and 2.6, respectively, compared with an expected value of 1.0. These observations, combined with the previously described coherence of 62%, indicate that the VA LV VFQ is a valid, but imprecise, functional assessment instrument. The fit statistics indicate that constructs other than visual ability contribute error to the measurement.

Test-Retest Reliability

Figure 8 is a scatterplot of item measures estimated from responses to the initial administration of the VA LV VFQ by 30 patients versus item measures estimated from responses to the re-administration of the VA LV VFQ to the same 30 patients 3 to 4 weeks later. The error bars represent ± 1 SE of the item measure estimate, and the diagonal line is the identity line. There is strong test-retest concordance with an ICC of 0.98 (95% confidence interval [CI]: 0.96–0.99). Figure 9 is a similar test-retest scatterplot for person measures. There is good test-retest concordance, with an ICC of 0.84 (95% CI: 0.69–0.92).

DISCUSSION

Our goal was to develop a self-report questionnaire to measure the difficulty visually impaired persons have in performing the daily living activities. This questionnaire could be used to tailor rehabilitation programs to meet individual patient needs and to measure outcomes of treatment strategies and rehabilitation programs. The 48-item VA LV VFQ was validated in both VA and private-sector programs, to facilitate comparison of outcomes from different service delivery settings and to increase the number of visually impaired female participants. Patients from multiple VA programs were included because the VA system provides separate rehabilitation programs for legally blind and partially sighted veterans, although there is some overlap in eligibility criteria for the programs.

The precision, validity, and reliability of the initial version of the VA LV VFQ were measured. The Andrich rating scale model was used to evaluate the required visual ability for each item and the threshold for each response category.^{23–25} The VA LV VFQ-48 includes five rating categories (*not difficult*, *slightly*

TABLE 5. Results of Rasch Analysis Applied to Rating of Difficulty for Each Activity

Entry No.	Raw Score	Count	Measure	Error	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Items
1	408	365	-2.6	0.16	0.91	-0.5	0.87	-0.5	Get dressed
2	393	365	-3.07	0.2	1.23	1.2	1.19	0.6	Keep clean
3	660	356	-0.1	0.07	1.17	2.3	1.25	2.7	Identify medicine
4	570	365	-0.71	0.08	1.11	1.3	1.27	2.3	Tell time
5	619	363	-0.38	0.08	0.67	-4.2	0.82	-1.8	Identify money
6	542	349	-0.71	0.08	0.92	-0.9	0.87	-1.1	Match clothes
7	507	361	-1.15	0.09	0.96	-0.4	1.18	1.3	Grooming
8	569	363	-0.69	0.08	0.75	-2.9	0.72	-2.4	Identify food
9	477	363	-1.47	0.1	0.8	-1.9	0.74	-1.6	Eat/drink neatly
10	434	344	-1.69	0.12	1.06	0.5	0.76	-1.3	Fix snack
11	527	298	-0.28	0.08	0.93	-0.8	0.85	-1.4	Prepare meals
12	658	357	-0.11	0.07	0.89	-1.5	0.92	-0.9	Use appliance dial
13	470	288	-0.59	0.09	1.08	0.9	1.3	2.4	Clean the house
14	669	321	0.31	0.07	1.25	3.4	1.26	2.9	Handle finances
15	752	336	0.57	0.07	1.05	0.8	1.05	0.6	Make out a check
16	554	358	-0.73	0.08	1.07	0.9	1.31	2.6	Take a message
17	743	360	0.29	0.07	0.64	-5.2	0.77	-2.7	Find something on a crowded shelf
18	556	354	-0.68	0.08	1.08	0.9	0.95	-0.4	Find public restroom
19	398	356	-2.62	0.16	1.03	0.2	1.43	1.6	Get around indoors in familiar places
20	500	352	-1.12	0.09	1.01	0.1	1.11	0.8	Get around outdoors in familiar places
21	704	336	0.34	0.07	0.61	-5.4	0.64	-4.1	Get around in unfamiliar places
22	711	307	0.66	0.07	1.2	2.7	1.37	4.5	Go out at night
23	654	344	-0.01	0.07	0.99	-0.1	1.14	1.5	Go down steps in dim light
24	713	359	0.14	0.07	1.14	2	1.22	2.5	Adjust to bright light
25	655	323	0.22	0.07	1.05	0.6	0.99	-0.1	Cross streets at traffic lights
26	489	205	0.85	0.09	1.63	7.1	1.77	7.8	Use public transportation
27	640	338	-0.02	0.07	1.22	2.9	1.25	2.7	Get around in a crowd
28	584	352	-0.48	0.08	1.26	3.2	1.39	3.6	Avoid bumping into things
29	647	359	-0.2	0.07	0.88	-1.6	0.86	-1.5	Recognize persons up close
30	842	323	1.1	0.07	0.98	-0.3	1	0.1	Recognize persons across room
31	898	360	0.96	0.07	0.81	-2.9	0.88	-1.6	Read street signs/store names
32	699	329	0.33	0.07	1.28	3.8	1.19	2.3	Read headlines
33	838	301	1.39	0.07	0.89	-1.4	0.86	-1.7	Read menus
34	923	321	1.54	0.07	0.95	-0.7	0.92	-1	Read newspaper/magazine articles
35	801	298	1.19	0.07	0.81	-2.6	0.76	-3	Read mail
36	952	319	1.72	0.07	1.01	0.1	0.96	-0.4	Read small print on package label
37	911	356	1.07	0.07	1.07	1.1	1.03	0.4	Read print on TV
38	870	351	0.95	0.07	1.04	0.5	0.99	-0.2	Keep place while reading
39	627	334	-0.05	0.07	0.78	-2.8	0.73	-2.8	Watch TV
40	520	214	0.77	0.09	0.97	-0.4	0.91	-0.9	Play table and card games
41	763	347	0.5	0.07	0.64	-5.2	0.63	-4.5	See photos
42	755	314	0.79	0.07	1.54	7.6	1.58	7.4	Work on favorite hobby
43	471	229	0.19	0.09	1.24	2.8	1.13	1.3	Go to movies
44	622	261	0.79	0.08	0.79	-2.7	0.75	-2.9	Go to spectator events
45	653	227	1.46	0.09	1.05	0.5	0.99	-0.1	Play sports
46	481	248	0.14	0.08	1.15	1.7	1.05	0.5	Do yard work
47	680	348	0.09	0.07	1.28	3.8	1.24	2.7	Sign your name
48	726	286	1.11	0.07	0.74	-3.6	0.72	-3.5	Read signs

difficult, moderately difficult, extremely difficult, and impossible). Analysis indicated that categories 2 (slightly difficult) and 3 (moderately difficult) performed poorly, even when the data from all five sites were combined. These responses were combined and the data reanalyzed with four rating categories rather than five. Further testing of the four rating categories (not difficult, slightly/moderately difficult, extremely difficult, and impossible) with an additional 50 subjects confirmed the validity of combining categories slightly difficult and moderately difficult. The use of four rating categories to measure the difficulty patients have in performing daily activities is consistent with reports from other studies.^{7,8}

A person-item map for the sample of 367 subjects (Fig. 5) was constructed from the measures of perceived visual ability for persons and the required visual ability to perform each of the 48 tasks. The map of persons and items for the total sample represented by all five sites demonstrates that the VA LV VFQ-48 best separates patients with visual ability in the range

of moderate to severe loss (+2 to -2 logits). The persons are distributed from most to least disabled, as would be expected from a sample of subjects with visual impairment that varies from no light perception to near normal vision. We also note that some patients did not have much difficulty performing any items. The person-item map demonstrates a floor effect. There are three activities (get dressed, keep clean, and get around indoors in places you know) that few patients reported difficulty performing. These items could be eliminated without any practical effect on the measurement range of the instrument. During development of the VA LV VFQ-48, clinicians and rehabilitation specialists indicated that the activities get dressed, keep clean, and get around in places you know are difficult for patients with vision loss. It is surprising that few patients reported difficulty with these items. One explanation for this finding is that rehabilitation professionals may see evidence of stains on clothing or assess hazards in the living environment, whereas the patients themselves may be unaware of the haz-

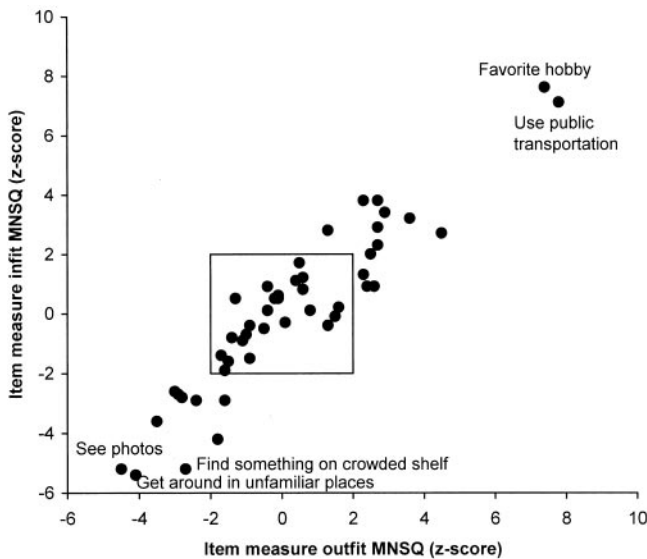


FIGURE 6. Scatterplot of the normalized mean square fit statistics for item measures, showing visual ability needed to perform each activity. The box includes infit and outfit mean square values that are within ± 2 SDs of the model's expectations.

ards or stains on their clothing. Difficulty getting dressed can be interpreted as physically dressing or more generally as coordinating clothing choices. Items are considered difficult on this questionnaire only if they are perceived by patients as related to vision loss. Difficulty dressing due to physical disabilities from stroke, arthritis, or other medical conditions would not be scored.

The person-item map is well centered on the BRC person measure distribution. The distributions for the sites other than the BRC extend to higher visual abilities than are covered by the items, indicating a ceiling effect. Fifteen percent of the patients had visual ability greater than 2 logits. The items selected are adequate to describe the functional ability of 85% of the subjects in the study. The more mildly impaired (those

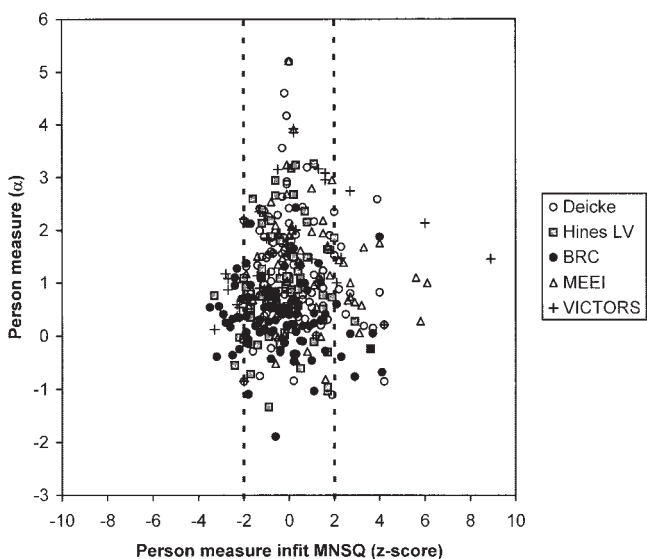


FIGURE 7. Scatterplot of the normalized mean square fit statistics for person measures, showing patients' abilities to perform activities. The box includes infit and outfit mean square values that are within ± 2 SDs of the model's expectations.

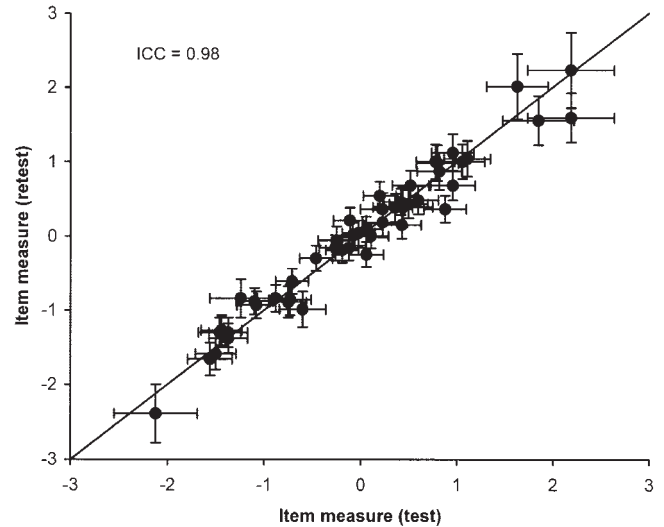


FIGURE 8. Scatterplot of item measures demonstrating test-retest concordance that was estimated from 30 patients initial responses to the VA LV VFQ-48 compared with the responses of the same patients 3 to 4 weeks later. Error bars, ± 1 SE of the item measure estimate; diagonal line: identity line.

with visual ability greater than 2 logits) are less well discriminated by the items.

We could consider adding more difficult items to the VA LV VFQ-48, such as reading signs at night or driving during poor weather conditions, to extend the scale and obtain more precise measurements of persons with higher levels of visual ability. However, discriminating differences among near-normal patients is not our goal in developing a low-vision visual functioning questionnaire. Our goals are to identify patients who need and will benefit from low-vision rehabilitation and to measure the outcomes of low-vision service delivery. It is a common problem that many patients with near-normal vision self-refer or are referred to low-vision clinics because they are dissatisfied with their refractive error correction or are experiencing normal changes in visual function associated with

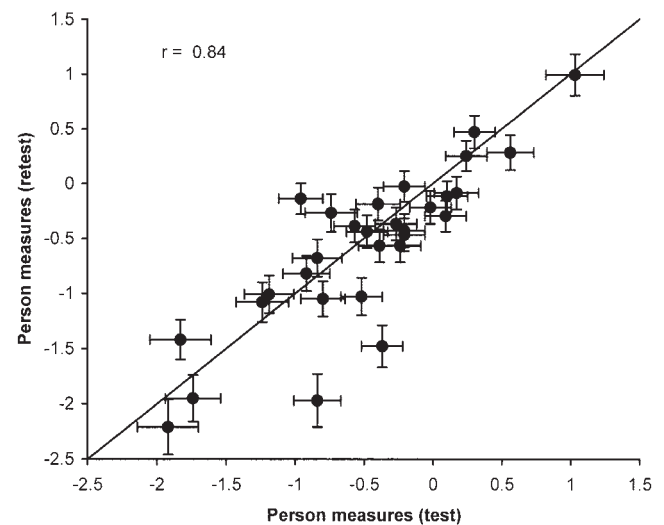


FIGURE 9. Scatterplot of person measures demonstrating test-retest concordance that was estimated from 30 patients initial responses to the VA LV VFQ-48 compared with the responses of the same patients 3 to 4 weeks later. Error bars, ± 1 SE of the item measure estimate; diagonal line: identity line.

aging, such as age-related changes in the lens, vitreous, and retina, as well as cellular alterations in the visual pathway and limbic system. Because eligibility and referral criteria differ between centers, some patients with near-normal vision receive services from some of the low-vision programs participating in this study.

The VA LV VFQ-48 does not include any driving items. Based on our experience with visual functioning questionnaires, there are considerable missing data on these items. Many older patients, especially those with moderate to severe vision loss self-limit driving or they may not be eligible to drive based on the criteria for licensing in the state of residence. Many of the patients with high visual ability may have self-referred to the low-vision clinic based on difficulty driving in daylight or at night. Difficulty with these tasks would not be reflected in the VA LV VFQ-48. The VA LV VFQ-48 would not be an appropriate instrument to evaluate the outcomes of a driving rehabilitation program. Specialized questionnaires to evaluate driving habits have been developed by other investigators.³²

The validity of the VA LV VFQ-48 was evaluated to determine the extent to which the pattern of patient responses agrees with the expectations of the model. Goodness of fit is represented by the transformed outfit and infit values for each item and person, presented as *z*-scores with an expected value of 0 and an SD of 1. Based on the scatterplot of the infit versus outfit mean square *z* scores for the 48 items (Fig. 6), residual errors for two items, *work on your favorite hobby* and *use public transportation* exceed the expected value by >5 SD. The item measure infit and outfit mean squares can be thought of as estimates of the variance of inherent difficulty within the sample of patients for each item. Large mean square values indicate high variance in the inherent difficulty distribution and small mean square values indicate low variance. Disagreements between the infit and outfit mean squares reflect a departure from symmetry of the inherent difficulty distribution for that item. High mean square values indicate imprecision in the item measure across persons.

The poor fit of these items is most likely due to the variability in the visual ability necessary to perform different hobbies or to access available forms of public transportation and the inherent difficulty of the task independent of vision loss. The visual demands (near vision) for stamp or coin collecting differ greatly from the visual demands and skills needed to play golf or table games (e.g., chess). Use of public transportation can include complex mobility tasks, such as traveling independently by subway or simply calling a cab to provide transportation to a specific location. Because there is so much variability in these items, they create noise and contribute little to the measurement properties of the instrument. They probably should not be used when estimating α .

Other items with both outfit and infit mean squares that exceed 2 SD include, identify medicine, handle finances, go out at night, get around in a crowd, avoid bumping into things, read headlines, and sign your name. These items may need to be rewritten to be more specific. As an example, patients may interpret "identify medicine" as meaning identify bottles or tubes of different medicines, identify shapes and colors of pills, or read print on the medicine bottle. Items that can be interpreted in different ways are often noisy. It is important to include the same items on the questionnaires administered before and after rehabilitation to assess sensitivity of the instrument to changes brought about by the rehabilitation program. Decisions to eliminate or rewrite items to refine the questionnaire should be deferred until after the postrehabilitation data collection and analysis.

Another explanation for misfit of some of the items is that the sample of subjects is largely composed of persons with

central visual acuity loss. Only 20% of the subjects in the initial analysis had diagnoses associated with peripheral visual field loss. Some of the items that misfit are those that expected to be sensitive to peripheral vision loss (e.g., get around in a crowd or avoid bumping into things). There is a basic assumption that the VA LV VFQ-48 is measuring one underlying construct, visual ability. Factor analysis may indicate that more than one construct is being measured. This possibility will be explored in future papers.

In some cases, there are multiple items representing the same level of difficulty. From a measurement perspective, some of these items could be omitted to shorten the instrument and decrease administration time and respondent burden. The prevailing practice in instrument development is to use the minimum number of items needed to measure a domain.^{13,14} However, when the instrument is used to measure low-vision outcomes, valuable information would be lost if clinically significant redundant items were eliminated. Our previous research with the NEI VFQ-25 indicates that item-specific changes occur as a result of rehabilitation.⁶ Low-vision rehabilitation enhances remaining vision for specific activities. As an example, read street signs and keep your place while reading represent the same ability level, but different low-vision devices would be used to make these activities easier to perform. Telescopic devices are indicated for distance activities, while magnifiers are used for near-reading activities. If we are measuring outcomes of service delivery, both items are useful. However, if we are measuring visual ability of patients undergoing cataract surgery, items could be based solely on the measurement properties of the scale, because cataract extraction restores visual ability. Changes on one item would reflect changes that would be expected on items of similar difficulty. Further modification of the VA LV VFQ-48 may be possible to reduce redundancy, eliminate or rewrite items with poor model fit to improve precision, and decrease the number of items on the instrument. If any items were to be dropped at this time, they would be the most misfitting ones and the ones easiest to perform by all the subjects. However, it is premature to edit the instrument before sensitivity of the items to change after rehabilitation can be evaluated. These results will be reported in future research reports.

Test-retest reliability was estimated for both persons and items, to confirm that we can discriminate people based on their estimated visual ability and items on the basis of their difficulty. Test-retest reliability yielded an ICC of 0.98 (95% CI: 0.96–0.99) for item measures and 0.84 (95% CI: 0.69–0.92) for person measures. These reliabilities can be compared with other instruments to measure visual impairment such as contrast sensitivity function tests although a true comparison should be done on the same patients. The VA LV VFQ-48 test-retest ICC for person measures is better than that reported for the VISTECH Chart with ICC ranges from 0.35 at low spatial frequencies to 0.93 at middle spatial frequencies and comparable to the Pelli-Robson CSF ICC of 0.86.³³

In conclusion, we have developed a valid and reliable tool to measure visual ability of low-vision patients with moderate to severe vision loss across diverse clinical settings. The instrument has been designed to assess activities that are addressed by rehabilitation so that it may have use as an outcomes measure. The items within the instrument were constructed to be concise, easily and universally understood, and with simple and consistent response categories. The psychometric properties of the instrument were evaluated using Rasch analysis. Based on results from the Rasch analysis, the scale has been modified to four response choices rather than five. Further refinement of the instrument is indicated after the data on sensitivity to change is evaluated. Comparison of pre- and postrehabilitation VA LV VFQ-48 scores will confirm that the

changes in values as a result of the intervention are resolvable with the instrument.

References

1. U.S. Department of Health and Human Services. *Vision Research A National Plan 1999-2003*: NIH Publication No. 98-4120. Bethesda, MD: National Eye Institute; 1998:117-130.
2. Department of Veterans Affairs, Veterans Health Administration, Blind Rehabilitation Service. *Coordinated Services for Blinded Veterans*. IB 11-59 (revised) P87250. Washington, DC: Department of Veterans Affairs; 1996:4.
3. Stelmack J. Quality of life of low-vision patients and outcomes of low-vision rehabilitation. *Optom Vis Sci*. 2001;78:335-342.
4. Massof RW, Rubin GS. Visual function assessment questionnaires. *Surv Ophthalmol*. 2001;45:531-548.
5. Massof RW. The measurement of vision disability. *Optom Vis Sci*. 2002;79:515-552.
6. Stelmack JA, Stelmack TR, Massof RW. Measuring low-vision outcomes with the NEI VFQ-25. *Invest Ophthalmol Vis Sci*. 2002;43:2859-2868.
7. 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.
8. 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.
9. Haymes SA, Johnston AW, Heyes AD. The development of the Melbourne Low-Vision ADL Index: A measure of visual disability. *Invest Ophthalmol Vis Sci*. 2001;42:1215-1225.
10. Gothwal VK, Lovie-Kitchin JE, Nutheti R. The development of the LV Prasad-Functional Vision Questionnaire: a measure of performance of visually impaired children. *Invest Ophthalmol Vis Sci*. 2003;44:4131-4139.
11. 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.
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. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alphas in the creation of minimum data sets. *Med Care*. 2004;42(suppl):1-17.
14. 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.
15. 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.
16. 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;119:1050-1059.
17. Steinberg EP, Tielsch JM, Schein OD, et al. The VF-14: an index of functional impairment in patients with cataract. *Arch Ophthalmol*. 1994;112:630-2638.
18. Stelmack J, Szlyk J, Joslin C, Swetland B, Myers L. Pilot study: use of the NEI VFQ-25 to measure outcomes of low vision rehabilitation services in the VA. In: Stuen C, Ardit A, Horowitz A, et al., eds. *Vision Rehabilitation: Assessment, Intervention and Outcomes*. Lisse, The Netherlands: Swets & Zeitlinger; 2000:774-776.
19. Stelmack J, Szlyk J, Stelmack T, et al. Use of the Rasch person item map in exploratory data analysis: a clinical perspective. *J Rehabil Res Dev*. 2004;41:233-242.
20. Szlyk JP, Stelmack J, Massof RW, et al. Performance of the veterans affairs low vision visual functioning questionnaire. *J Vis Impair Blind*. 2004;98:261-275.
21. Stelmack J, Rosenbloom A, Brennehan C, Stelmack T. Patients' perceptions of the need for low vision devices. *J Vis Impair Blind*. 2003;97:521-535.
22. Dunn CM, Chadwick G. Code of Federal Regulations. In: Allen W, ed. *Protecting Study Volunteers in Research*. Boston, MA: Center Watch, Inc.; 1999:203.
23. Wright BD, Masters GN. *Rating Scale Analysis*. Chicago, IL: MESA Press; 1982.
24. Andrich D. A Rating formulation for ordered response categories. *Psychometrika*. 1978;43:561-573.
25. Linacre JM, Wright B. *A User's Guide to Winsteps Rasch-Model Computer Program*. Chicago, IL: MESA Press; 2001.
26. Luce RD, Tukey JW. Simultaneous conjoint measurement: a new type of fundamental measurement. *J Math Psychol*. 1964;1:1-27.
27. Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. *Stat Med*. 1994;13:2465-2476.
28. Menzel H. A new coefficient for scalogram analysis. *Public Opin Q*. 1953;17:268-280.
29. Chakravarti IM, Laha RG, Roy J. *Handbook of Methods of Applied Statistics*. Vol. 1. New York: John Wiley & Sons; 1967:392-394.
30. Bond TG, Fox CM. *Applying the Rasch Model*. Mahwah, NJ: Erlbaum; 2001.
31. Smith RM. The distributional properties of Rasch item fit statistics. *Educ Psychol Meas*. 1991;51:541-565.
32. Owsley C, Stalvey B, Wells J, Sloane ME. Older drivers and cataract: driving habits and crash risk. *J Gerontology A: Biological Sciences Medical Sciences*. 1999;54:203-211.
33. Rubin GS. Reliability and sensitivity of clinical contrast sensitivity tests. *Clin Vision Sci*. 1988;2:169-177.