Measuring Outcomes of Vision Rehabilitation with the Veterans Affairs Low Vision Visual Functioning Questionnaire

Joan A. Stelmack,[1,2,3] *Janet P. Szlyk,*[4,3] *Thomas R. Stelmack,*[4,2,3] *Paulette Demers-Turco,*[5] *R. Tracy Williams,*[6,7,2] *D'Anna Moran,*[1] *and Robert W. Massof*[8]

Purpose. To evaluate the sensitivity to change, in patients who undergo vision rehabilitation, of the Veteran Affairs (VA) Low Vision Visual Functioning Questionnaire (LV VFQ-48), which was designed to measure the difficulty visually impaired persons have in performing daily activities and to evaluate vision rehabilitation outcomes.

Methods. Before and after rehabilitation, the VA LV VFQ-48 was administered by telephone interview to subjects from five sites in the VA and private sector. Visual acuity of these subjects ranged from near normal to total blindness.

Results. The VA LV VFQ exhibited significant differential item functioning (DIF) for 7 of 48 items (two mobility tasks, four reading tasks, and one distance-vision task). However, the DIF was small relative to baseline changes in item difficulty for all items. Therefore, the data were reanalyzed with the constraint that item difficulties do not change with rehabilitation, which assigns all changes to the person measure. Subjects in the inpatient Blind Rehabilitation Center (BRC) program showed the largest changes in person measures after vision rehabilitation (effect size = 1.9; *t*-test $P < 0.0001$). The subjects in the outpatient programs exhibited smaller changes in person measures after rehabilitation (effect size = 0.29; *t*-test $P < 0.01$). There was no significant change in person measures for the control group (test–retest before rehabilitation).

Conclusions. In addition to being a valid and reliable measure of visual ability, the VA LV VFQ-48 is a sensitive measure of changes that occur in visual ability as a result of vision rehabilitation. Patients' self-reports of the difficulty they experience performing daily activities measured with this instrument can be used to compute a single number, the person measure that can serve as an outcome measure in clinical studies. The VA LV VFQ-48 can be used to compare programs that offer different levels of intervention and serve patients across the continuum of vision loss. (*Invest Ophthalmol Vis Sci.* 2006;47:3253–3261) DOI:10.1167/iovs.05-1319

There is a growing consensus that the effects of low-vision rehabilitation must be measured in terms of the patient's self-reported ability to perform daily activities.[1–3] In the few vision rehabilitation outcome studies that involved rating-scale questionnaires, such as the 25-item National Eye Institute Visual Functioning Questionnaire (NEI VFQ-25), small improvements after intervention were reported.[4–6] However, in all such studies, sums of ratings across items (i.e., Likert scores) were used as the summary measure of each person's functional ability or quality of life.

Aside from the limitations of Likert scores as measures,[7] Likert scoring of rating scale questionnaires can underestimate the effects of rehabilitation if there are nonresponsive items.[8] In other words, if a questionnaire contains items that are not relevant to the rehabilitation efforts, then one would expect patient post-rehabilitation responses to be unchanged from pre-rehabilitation responses. For example, if there are items that ask the patient to rate the difficulty of mobility tasks intermingled with items that ask about the difficulty of reading tasks and the intervention is limited to stand magnifiers, then one would expect changes in responses to the reading tasks only. From this example, we can see that the conclusions drawn about the effectiveness of rehabilitation depend on the choice of items in the questionnaire.

This type of differential item-functioning with respect to the effects of rehabilitation was observed in an earlier vision rehabilitation outcome study of two Veterans Affairs (VA) rehabilitation programs. Using Rasch analysis with the NEI VFQ-25 (plus supplement), that earlier study demonstrated that the change in difficulty of some items after rehabilitation was more than the change for others.[8] Only 7 of the 34 items tested were sensitive to change after rehabilitation: reading ordinary print in newspapers (item 5); going out to see movies, plays, or sports events (item 14); reading small print in a telephone book, on a medicine bottle, or on a legal form (item A3); figuring out whether bills are accurate (item A4); see well up close (item 6); reading street signs or the names of stores (item 8); and seeing and enjoying programs on television (item A8). Only 4 of these (items 5, 14, A3, and A4) were sensitive to change after rehabilitation in both rehabilitation programs. When pre-rehabilitation item ordering was compared to post-rehabilitation item ordering, disagreement was observed because of those items that were selectively responsive to rehabilitation.

In response to the measurement shortcomings of some instruments, the Veteran Affairs Low Vision Visual Functioning Questionnaire (VA LV VFQ-48) was developed to serve as a

From the [1]Blind Rehabilitation Center, Edward Hines VA Hospital, Hines, Illinois; the [4]Jessie Brown VAMC, Chicago, Illinois; [5]Low Vision Optometry Practice, North Andover, Massachusetts; the [6]Lois and Edwin Deicke Center for Vision Rehabilitation, Wheaton, Illinois; the [2]Illinois College of Optometry, Chicago, Illinois; the [3]Department of Ophthalmology and Visual Science, University of Illinois at Chicago College of Medicine, Chicago, Illinois; the [7]Department of Ophthalmology, Loyola University School of Medicine, Maywood, Illinois; and the [8]Wilmer Eye Institute, Johns Hopkins University School of Medicine, 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 National Eye Institute Grant EY-012045 (RWM).

Submitted for publication October 6, 2005; revised March 8, 2006; accepted May 25, 2006.

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.

patient evaluation tool and an outcome measure for vision rehabilitation.[9–12] Items were chosen to reflect activities that patients with visual impairment report are important to them and are addressed by rehabilitation programs. The VA LV VFQ-48 is administered by telephone to record patients' self-reports of the difficulty they experience performing daily activities. The instrument was developed by a team of clinicians, rehabilitation specialists, scientists, research staff, and persons with vision loss using a modified Delphi Method and validated by Rasch analysis using the Andrich Rating Scale Model. Development of the VA LV VFQ-48 and its psychometric properties are described in prior publications.[9–12]

The advantage of using the Andrich Rating Scale Model is that responses to the items on the questionnaire can be interpreted in terms of latent person and item variables that are the theoretical objects of measurement.[13–22] Person and item measures have equal status in the model. Theoretically, the difference between person and item measures is functional reserve. Functional reserve is defined as the difference between the visual ability a patient possesses and the visual ability needed to perform the activity.[23,24] The visual ability of each patient is dependent on the extent and type of visual impairment. A patient finds an activity easy to perform when his or her visual ability exceeds that necessary to perform the activity. When visual ability is only slightly greater than what the activity requires, a patient's functional reserve is limited, and the activity is reported to be more difficult. When visual ability is less than that required by the activity, functional reserve is negative and the patient is expected to report that he or she cannot perform the activity at all. The difficulty rating selected by the patient when responding to the VA LV VFQ-48 question, "Is it difficult to. . . ?" is an estimate of functional reserve. The patient judges each activity to be not difficult, slightly or moderately difficult, extremely difficult, or impossible based on the functional reserve available before rehabilitation and the increase in functional reserve due to the effects of assistive devices and techniques to maximize residual vision and/or substitute for visual information that are expected to make activities easier to perform after vision rehabilitation. Changes in functional reserve may occur on an item-by-item basis.

The change in functional reserve after rehabilitation can be assigned as a change in either the person or item measure, or both. There are four methods to parse the change into its components: (1) Pre- and post-rehabilitation item measures can be anchored to the items that exhibit the smallest change, the average of which is then transferred to post-rehabilitation person measures. All other changes are assigned to the items and/or persons. This technique was used for analysis of the pre and post low-vision-rehabilitation ratings on the NEI VFQ-25.[8] The small 0.3 logit (log odds) change in the anchor items was assigned to the person measures; the rest of the change in post-rehabilitation appeared in the items. (2) Rasch analysis is performed separately for pre- and post-rehabilitation ratings and the pre- to post-rehabilitation scales are compared by normalizing all measures to the average item measure change from pre- to post-rehabilitation. In the NEI VFQ-25 data previously reported, 7 of 39 items stood out as exhibiting the largest change. The rest of the items fell on a regression line parallel to the identity line. For this reason, method 1 was used, assigning the smallest change to the person measures and rescaling all the items. If method 2 had been used in the analysis of the NEI VFQ-25 data, the average effect of rehabilitation would have been very small, because the many items that did not change would have diluted the effects of intervention.

The last two methods, which were used in the present study, involve "racking" and "stacking."[25,26] (3) Racking refers to aligning the pre- and post-rehabilitation items side by side in the data matrix, as if they are different items. This method assumes that the person measure does not change as a result of the intervention and that all the change between pre- and post-rehabilitation occurs in the item measures. (4) The fourth method is to stack the pre- and post-rehabilitation ratings in the data matrix as if there are two different groups of respondents (stacking technique). This method assumes that no change occurs in the item measures and that all the change from pre- to post-rehabilitation occurs in the person measures.

We already have established that the VA LV VFQ-48 is a valid and reliable measure of visual ability for a sample of legally blind patients seeking services in an inpatient VA Blind Rehabilitation Center (BRC) and in a variety of outpatient low-vision rehabilitation programs.[9–12] The present paper takes the next step in validating the VA LV VFQ-48 as an outcome measure by evaluating its sensitivity and differential item functioning to changes that result from vision rehabilitation. Secondarily, we offer estimates of effect sizes from changes in VA LV VFQ-48 measures for different types of vision rehabilitation programs.

Methods

Instruments

The VA LV VFQ-48 includes four questions. Question 1, "Is it difficult to . . . ?" is asked about all 48 items. Response choices include: *not difficult*, *slightly/moderately difficult*, *extremely difficult*, *impossible*, and *do not do it for nonvisual reasons* (which is scored as missing data). Question 2 is asked only if it is difficult to perform the activity. "Is it because of your vision?" has two response choices: *yes* and *no*. If the activity is difficult to perform because of vision loss, question 3, "Do you want training. . . ?" is asked. Response choices for question 3 are *yes* and *no*. Questions 4 asks, "How do you usually. . . ?" Response choices are *own eyes* or *eye glasses*, *vision devices/techniques*, *other senses/nonvisual devices*, *someone helps me*, and *not applicable*. The present paper is concerned only with the responses to question 1.

The 48 items on the VA LV VFQ are presented in the Results section.

Subjects

Consecutive patients were recruited from five clinical sites from May 2000 to May 2002. Sites included Edward Hines Veterans Affairs Hospital Blind Rehabilitation Center (Hines BRC), Hines VA Outpatient Low Vision Rehabilitation/Eye Clinic, Jessie Brown Veterans Affairs 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). The VA sites include a suburban hospital (Hines) and an inner-city medical center (Jessie Brown). The private sector sites include a not-for-profit agency (Deicke Center) located in the Chicago suburbs and a hospital-based rehabilitation service located in a major city (MEEI). The Deicke Center, MEEI, and VICTORS are outpatient low-vision rehabilitation programs that serve a representative sample of patients with low vision who are referred or self-refer to low-vision clinics. These programs primarily provide low-vision evaluation, prescription of low-vision devices, and instruction in their use (average of two therapy sessions for Deicke Center and MEEI, four therapy sessions for VICTORS). The Hines BRC is a comprehensive inpatient program that serves only legally blind veterans. The multidisciplinary team at the Hines BRC provides evaluation, psychosocial intervention, recreation, and classes in orientation and mobility, living skills, manual skills and visual skills (a mean of 40 days' hospital stay). A control group consisted of subjects on the waiting list for services at Hines VA who had not participated in low-vision services or comprehensive blind rehabilitation. The control group was used to validate stability of the measurement over time. Because the population of patients at the Hines BRC is more severely visually impaired, and the program is more comprehensive the outcomes for the outpatient low-vision rehabilitation pro-

TABLE 1. Habitual Visual Acuity of Subjects

Site Name	Inpatient BRC	Outpatient LV Programs	Control Group	Total
Normal VA				
(>20/40)*	0/139	8/116	0/30	8/285
Log MAR <0.301 (%)†	0	7	0	3
SD Log MAR	—	0.069	—	0.069
Near-normal impairment				
(20/40–20/60)*	6/139	39/116	4/30	49/285
0.301–0.477 (%)†	4	34	13	17
SD Log MAR	0.019	0.075	0.086	0.074
Moderate impairment				
(20/70–20/160)*	35/139	40/116	4/30	79/285
0.544–0.903 (%)†	25	34	13	28
SD Log MAR	0.135	0.137	0.082	0.139
Severe impairment				
(20/200–20/400)*	83/139	25/116	16/30	124/285
1.000–1.301 (%)†	60	22	54	43
SD Log MAR	0.111	0.103	0.139	0.113
Profound impairment				
(20/500–20/1000)*	11/139	4/116	4/30	19/285
1.398–1.698 (%)†	8	3	13	7
SD Log MAR	0.068	0.127	0.056	0.120
Near-total impairment/total impairment (NLP)				
(<20/1000)*	4/139	0/116	2/30	6/285
Log MAR >1.698 (%)†	3	0	7	2

Inpatient BRC: Hines Blind Rehabilitation Center, Edward Hines VA Hospital; Outpatient low-vision programs: VICTORS (Visual Impairment Center to Optimize Remaining Sight), Deicke Center for Visual Rehabilitation, and the Massachusetts Eye and Ear Infirmary. Control group: Edward Hines VA Hospital. Data are Based on evaluation of best eye by ICD-9-CM classification. NLP, No light perception.

* Snellen VA.

† Log MAR VA.

grams and the inpatient blind rehabilitation center program are reported separately.

The analysis of the VA LV VFQ-48's sensitivity to change after vision rehabilitation is based on data from 285 subjects (139 subjects from Hines BRC, 34 subjects from VICTORS, 54 subjects from Deicke, 28 subjects from MEEI, and 30 control subjects from the Hines Outpatient Low Vision Rehabilitation/Eye clinic who participated only in the test–retest evaluation). The VA LV VFQ-48 was administered by telephone before rehabilitation and 3 months after rehabilitation. A total of 116 subjects participated from the outpatient low-vision programs and 136 subjects from the inpatient Hines BRC program.

The inpatient BRC and control group patients were 93% male; whereas, the outpatient low-vision program patients were 62% male. A higher proportion of patients were male due to the veterans participating in both groups. Most older veterans are male due to recruitment into the armed forces during World War II, the Korean Conflict, and the Vietnam War. The mean age of the patients was 71 years in the inpatient BRC group (SD 11.3; range, 42–87), 73 years in the control group (SD 9.4; range, 49–88), and 74 years in the outpatient low-vision programs (SD 12.4; range, 27–90). The most frequent eye diagnoses were macular degeneration (69% inpatient BRC, 50% control group, 67% outpatient low-vision rehabilitation programs), diabetic retinopathy (11% inpatient, 27% control, 14% outpatient), glaucoma (8% inpatient, 8% control, 7% outpatient) and neurologic disorders (5% inpatient, 5% control, 7% outpatient). The habitual visual acuity of subjects is described in Table 1.

Patients who participated were required to complete the questionnaire independently. Surrogates (e.g., family members) were not allowed to assist or complete the questionnaire, to ensure that the responses were self-reports based on the patient's perception of difficulty performing activities. Eight patients with serious health conditions (e.g., major depression, terminal illness) where issues of respondent burden were a concern were excluded by clinicians at each site, using medical information that was available at the site. Clinicians at the sites also excluded 11 patients with severe cognitive loss who could not reliably self-report the difficulty they experienced with daily activities. The study was conducted in compliance with the tenets of the Declaration of Helsinki for research in human subjects.[27] Informed consent was obtained according to the protocol requirement of the local institutional review board (IRB) for each facility, with oversight from the IRBs at Hines Hospital and the University of Illinois at Chicago.

Two trained interviewers located at the Department of Ophthalmology, University of Illinois at Chicago, conducted interviews for the

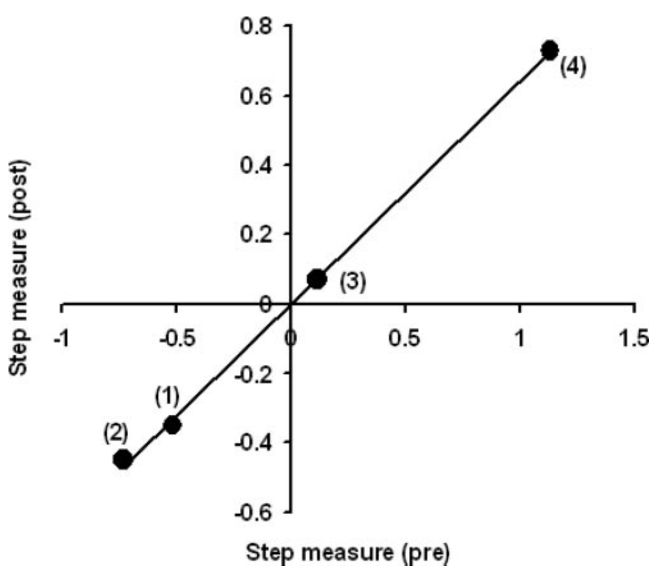

FIGURE 1. Estimated response category thresholds for post-rehabilitation data (step measures) were nearly identical with estimated response category thresholds for pre-rehabilitation data.

Table 2. Estimated Item Measures and Standard Errors of the Estimate

Items	Isolated Pre-rehab Measure	SE	Isolated Post-rehab Measure	SE	Racked Pre-rehab Measure	SE	Racked Post-Rehab Measure	SE	Stacked Pre/Post-rehab Measure	SE	Isolated Pre-rehab Infit MNSQ	Stacked Pre/Post Infit MNSQ	MNSQ Ratio
Physically get dressed	−2.44	0.17	−1.66	0.16	−2.64	0.22	−2.38	0.19	−2.74	0.15	1.37	1.3	1.05
Keep clean/keep your clothes clean	−1.28	0.09	−0.96	0.11	−1.27	0.13	−1.73	0.15	−1.64	0.1	1.68	1.56	1.08
ID Medicine	−0.1	0.06	−0.24	0.08	0.31	0.08	−0.65	0.11	−0.18	0.07	1.28	1.21	1.06
Tell time	−0.57	0.07	−0.88	0.1	−0.27	0.09	−1.58	0.14	−0.92	0.08	1.18	1.15	1.03
Identify money	−0.33	0.06	−0.26	0.08	0	0.09	−0.67	0.11	−0.39	0.07	0.82	0.8	1.03
Match clothes	−0.65	0.07	−0.29	0.08	−0.3	0.1	−0.77	0.11	−0.62	0.08	0.97	1.06	0.92
Groom self	−1.08	0.08	−0.83	0.1	−0.84	0.11	−1.41	0.13	−1.25	0.09	0.99	0.96	1.03
Identify food on a plate	−0.57	0.07	−0.18	0.07	−0.16	0.09	−0.57	0.1	−0.43	0.07	0.81	0.81	1.00
Eat and drink neatly	−1.31	0.09	−0.71	0.1	−1.14	0.12	−1.26	0.13	−1.36	0.09	0.85	0.86	0.99
Fix a snack	−1.34	0.09	−1.58	0.16	−1.28	0.13	−2.48	0.2	−1.93	0.11	1.14	1.21	0.94
Prepare meals	−0.23	0.07	−0.3	0.09	0.16	0.09	−0.84	0.12	−0.35	0.08	1.02	1.0	1.02
Use appliance dials	−0.14	0.06	−0.29	0.08	0.31	0.08	−0.8	0.11	−0.23	0.07	0.94	0.92	1.02
Clean the house	−0.54	0.07	−0.38	0.09	−0.18	0.1	−0.88	0.12	−0.59	0.08	1.06	1.06	1.00
Handle finances	0.34	0.06	0	0.07	0.8	0.08	−0.39	0.1	0.26	0.06	1.35	1.17	1.15
Make out a check	0.55	0.06	0.08	0.07	1.01	0.08	−0.25	0.1	0.46	0.06	1.16	0.98	1.18
Take a message	−0.45	0.07	−0.58	0.09	−0.18	0.09	−1.14	0.12	−0.7	0.08	1.07	1.03	1.04
Find something on a crowded shelf	0.31	0.06	0.29	0.06	0.71	0.08	0.03	0.09	0.39	0.06	0.76	0.72	1.06
Find public restroom	−0.57	0.07	−0.22	0.08	−0.29	0.09	−0.7	0.11	−0.58	0.07	1.22	1.19	1.03
Get around indoors in places you know	−2.15	0.14	−2.4	0.25	−2.26	0.19	−3.24	0.28	−2.88	0.16	1.25	1.19	1.05
Get around outdoors in places you know	−0.99	0.08	−0.95	0.11	−0.78	0.11	−1.55	0.14	−1.28	0.09	0.98	0.96	1.02
Get around in unfamiliar places	0.26	0.06	0.41	0.06	0.69	0.08	0.15	0.09	0.44	0.06	0.68	0.67	1.01
Go out at night	0.37	0.06	0.71	0.06	0.85	0.08	0.48	0.09	0.66	0.06	1.22	1.27	0.96
Get down steps in dim light	−0.17	0.06	0.14	0.07	0.21	0.08	−0.2	0.09	−0.05	0.07	1.12	1.11	1.01
Adjust to bright light	0.11	0.06	0.34	0.06	0.55	0.08	0.03	0.09	0.28	0.06	1.28	1.27	1.01
Cross streets at traffic light	−0.08	0.07	−0.13	0.08	0.29	0.09	−0.63	0.11	−0.19	0.07	1.02	1.07	0.95
Use public transportation	−0.18	0.09	−0.21	0.12	0.29	0.12	−0.63	0.17	−0.19	0.1	1.1	1.06	1.04
Get around in a crowd	−0.36	0.07	−0.06	0.07	−0.02	0.09	−0.49	0.1	−0.33	0.07	0.9	0.89	1.01
Avoid bumping into things	−0.76	0.07	−0.5	0.09	−0.48	0.1	−0.99	0.12	−0.84	0.08	0.96	0.95	1.01
Recognize persons up close	−0.34	0.06	0.11	0.07	0.01	0.09	−0.28	0.09	−0.2	0.07	1.1	1.23	0.89
Recognize persons across the room	1.06	0.06	1.4	0.06	1.55	0.07	1.31	0.07	1.57	0.05	0.96	1.06	0.91
Read street signs	0.97	0.06	0.82	0.06	1.43	0.07	0.68	0.08	1.14	0.06	0.9	0.88	1.02
Read headlines	0.07	0.06	−0.34	0.08	0.63	0.08	−0.86	0.11	−0.04	0.07	1.2	1.1	1.09
Read menus	1.31	0.06	1.17	0.06	1.83	0.07	1.11	0.07	1.62	0.06	0.78	0.88	0.89
Read newspapers/magazines	1.4	0.07	0.54	0.06	1.93	0.07	0.38	0.08	1.28	0.06	0.89	0.96	0.93
Read mail	0.94	0.06	0.16	0.07	1.45	0.07	−0.18	0.09	0.77	0.06	0.87	0.84	1.04
Read small print on packages	1.65	0.07	0.82	0.06	2.14	0.08	0.69	0.08	1.56	0.05	0.9	0.96	0.94
Read print on TV	1.28	0.06	1.3	0.06	1.76	0.07	1.24	0.07	1.66	0.06	0.81	1.0	0.81
Keep your place while reading	0.91	0.06	0.38	0.06	1.4	0.07	0.13	0.09	0.86	0.06	1.16	0.96	1.21
Watch TV	0.05	0.06	0.14	0.07	0.53	0.08	−0.24	0.09	0.14	0.06	0.89	0.84	1.06
Play table and card games	0.63	0.07	0.64	0.07	1.21	0.08	0.51	0.1	0.88	0.07	0.85	0.85	1.00
See photos	0.57	0.06	0.22	0.06	0.99	0.07	−0.09	0.09	0.51	0.06	0.62	0.71	0.87
Work on your favorite hobby	0.88	0.07	0.78	0.06	1.43	0.08	0.64	0.09	1.12	0.06	1.36	1.31	1.04
Go to movies	0.55	0.08	0.64	0.08	1.07	0.1	0.51	0.11	0.79	0.08	0.96	1.12	0.86
Go to spectator events	0.65	0.07	0.78	0.07	1.12	0.09	0.56	0.1	0.89	0.07	0.86	1.01	0.85
Play sports	1.27	0.09	1.57	0.09	1.88	0.1	1.61	0.12	1.86	0.08	1.26	1.63	0.77
Do yard work	0.11	0.07	0.24	0.08	0.64	0.1	−0.06	0.11	0.28	0.08	1.14	1.19	0.96
Sign your name	−0.4	0.07	−0.49	0.09	0.01	0.09	−1.04	0.12	−0.53	0.07	0.87	0.86	1.01
Read store signs	0.78	0.06	0.73	0.06	1.23	0.08	0.57	0.08	0.99	0.06	0.73	0.83	0.88

Estimates were derived from Rasch analyses performed on pre-rehabilitation data alone (Isolated Pre-rehab Measure), post-rehabilitation data alone (Isolated Post-rehab Measure), racked data for pre-rehabilitation items (Racked Pre-rehab Measure), racked data for post-rehabilitation items (Racked Post-rehab Measure), and stacked data for pre- and post-rehabilitation responses combined (Stacked Pre/Post-rehab Measure). The last three columns list the infit mean squares for the prerehabilitation data alone (Isolated Pre-rehab Infit MNSQ) and for the stacked data for pre- and post-rehabilitation combined (Stacked Pre/Post Infit MNSQ) and the ratio of these two mean squares (MNSQ Ratio).

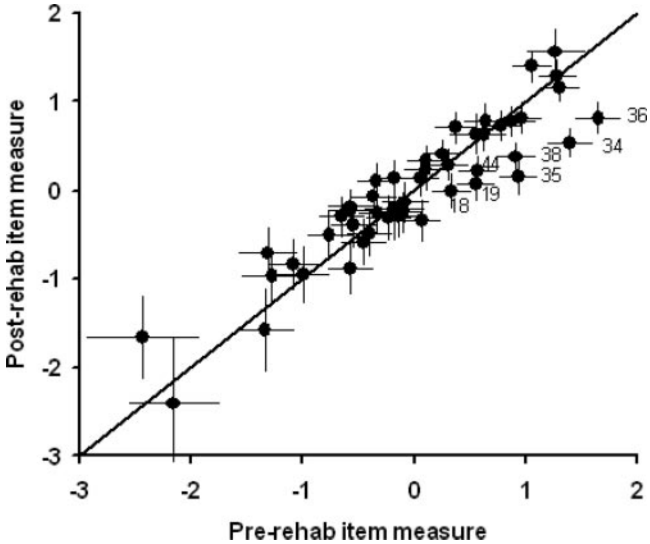

FIGURE 2. The scatterplot of pre- and post-rehabilitation item measures from separate Rasch analyses demonstrated differential item functioning for seven items (labeled with item numbers). Error bars, ±3 SE.

VICTORS patients. Three trained interviewers located at the Hines BRC conducted interviews for the other sites. A written script was used in administration. All interviewers were trained to administer the VA LV VFQ-48. An experienced interviewer also scored the instrument during training of new interviewers and periodically as a quality-control measure to confirm agreement in scoring. Administration time for the VA LV VFQ-48 (all four questions) varied from 25 to 35 minutes.

Analysis

Rasch analysis[21,22] with the Andrich Rating Scale Model was run separately for the pre- and post-rehabilitation item difficulty ratings of the 285 subjects.[28] Rasch analyses were repeated with pre- and post-rehabilitation data racked as one group of persons responding to two different sets of items, and, again, with pre- and post-rehabilitation data stacked as two different groups of persons responding to one set of items. For each type of analysis, Winsteps provided estimates of person measures, item measures, and response category thresholds.[29] It also provided estimates of mean square fit statistics for each person and item and estimates of separation reliabilities for distributions of person and item measures. Because all estimates are based on interval scales and are approximately normally distributed, parametric tests, and Pearson correlations were used in the statistical analyses.

Because the effects of vision rehabilitation can be item-specific, we examined changes in item measures from before to after rehabilitation. Within classic test theory, changes in item difficulty between populations (i.e., conditional probability of a correct answer) is called DIF. In the present study, the two groups under consideration consisted of the same patients, but before and after rehabilitation. If the person's vision were changed by the intervention, we would expect only a change in the person's ability and no change in item measures. In this example there would be no evidence of DIF. However, low-vision devices and adaptive training can make the activity described by an item more or less difficult without affecting the difficulty of other activities. For example, a magnifier could make it easier to read newspapers and magazines (item 34), but have no effect on the person's ability to get down steps in dim light (item 23). In this example, the use of a magnifier causes DIF. To look for evidence of DIF, we statistically compared changes in item measures from before to after rehabilitation. A t statistic was estimated for each item by taking the difference between pre- and post-rehabilitation item measures and dividing by the square-root of the sum of squares of the standard errors of the estimate (degrees of freedom determined by the number of respondents to each item). The α level was corrected for multiple item comparisons by using the Bonferroni method.

RESULTS

In our earlier validation study of the VA LV VFQ-48,[9] we observed disordering of the response category thresholds, with underutilization of two of the response categories. As illustrated by the scatter plot in Figure 1, the estimated response category thresholds for post-rehabilitation data (called "step measures" in Winsteps) were nearly identical with estimated response category thresholds for pre-rehabilitation data.

The same disordering of category thresholds was seen for both the pre- and post-rehabilitation data (i.e., the category 2 threshold preceded the category 1 threshold). Consequently, as before, we combined the responses for categories 2 and 3 and reanalyzed the data. All results presented in this article are based on analyses using the recoded response categories.

Evaluation of DIF

Table 2 lists each of the items in the VA LV VFQ-48 and the item measures and standard errors estimated from Rasch analyses of pre-rehabilitation data alone, post-rehabilitation data alone, racked pre-rehabilitation data, racked post-rehabilitation data, and stacked pre- and post-rehabilitation data. Also included are the infit mean squares for each item for the pre-rehabilitation data alone, and for the stacked pre- and post-rehabilitation data.

DIF occurs if the post-rehabilitation item measure is significantly different from the pre-rehabilitation item measure.[21,30]

Figure 2 illustrates a scatterplot of pre- versus post-rehabilitation item measures when Rasch analysis was performed separately on the two sets of data (isolated measures in Table 2).

The error bars represent 99% confidence intervals. The seven points labeled with item numbers exhibited significant DIF. That is, after correction for multiple comparisons, these items required significantly less ability (i.e., easier) after rehabilitation than before rehabilitation ($P < 0.05$ with Bonferroni

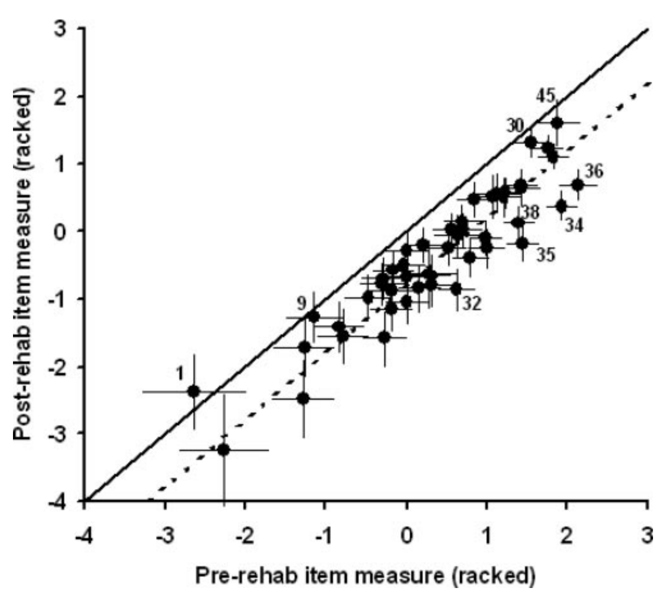

FIGURE 3. The scatter plot of pre- and post-rehabilitation item measures was estimated from analysis of racked data. Forty-four of the 48 items (92%) fell below the identity line, but on a parallel trend line. The post-rehabilitation item measure was, on average, 0.8 logit less than the pre-rehabilitation item measure (*dashed line*).

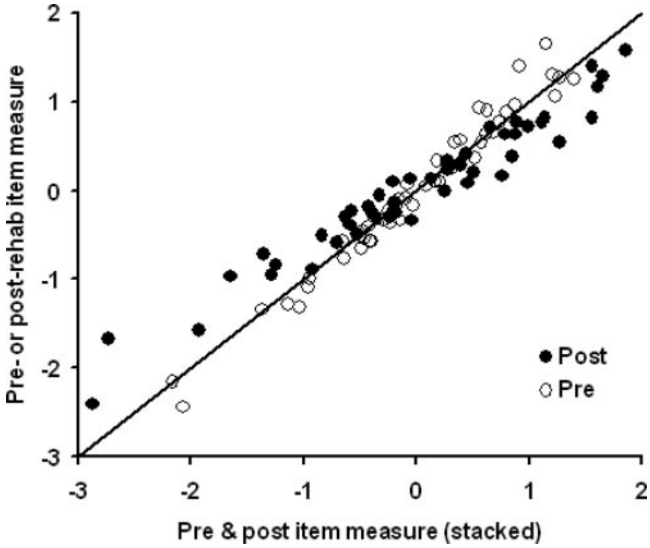

FIGURE 4. The scatterplot demonstrates that item measures from the Rasch analysis of stacked data agreed with both pre- and post-rehabilitation item measures.

correction). Items 18 and 19 are mobility tasks, items 34 to 36 and 38 are reading tasks, and item 44 is a distance vision task.

When the analysis is performed on stacked pre- and post-rehabilitation data, DIF would be manifested by greater infit mean squares relative to mean squares estimated from pre-rehabilitation data alone. This manifestation of DIF in the mean squares is expected because the pre-rehabilitation ratings will be more difficult than the post-rehabilitation ratings for those items with DIF, giving rise to greater response variability relative to the response expected by the model. The last column of Table 2 lists the ratios of the infit mean squares for the two sets of item measures (analogous to F ratios). The ratios are not significantly different from 1.0 for most items. The items that exhibited DIF in Figure 2, did not show any evidence of DIF in the ratio of mean squares. The only items that suggested the possibility of DIF in the analysis of stacked data were read print on TV (item 37) and play sports, (item 45).

When the analysis is performed on racked pre- and post-rehabilitation data, DIF should be exaggerated, because all differences between pre- and post-rehabilitation responses are assigned to the item measures. Figure 3 illustrates a scatterplot of pre- versus post-rehabilitation item measures estimated from an analysis of the racked data. Forty-four of the 48 items (92%) fall below the identity line, but on a parallel trend line. On average, for these items the post-rehabilitation item measure is 0.8 logit less than the pre-rehabilitation item measure (dashed line). The pre-rehabilitation item measures were not significantly different from the post-rehabilitation item measures for four of the items (physically get dressed [item 1]; eat and drink neatly [9]; recognize people across the room [30]; and play sports [45]). Five items (read headlines [32]; read newspapers and magazines [34]; read mail [35]; read small print on packages [36]; and keep place while reading [38]) were significantly less difficult post-rehabilitation than would be expected from the average change in item measure from pre- to post-rehabilitation (relative to the dashed line). All these items, except item 32, exhibited DIF when separate analyses were performed on pre- and post-rehabilitation data (Fig. 2).

For the five items that were significantly different from the average, the magnitude of DIF relative to the overall change in item measures from pre- to post-rehabilitation ranges from 0.47 to 0.83 logit. Although there is evidence of rehabilitation-dependent DIF in the VA LV VFQ-48, particularly for reading items, the main effect of rehabilitation on item measures is a constant decrease in required visual ability across all but four items, which can be interpreted as a constant change in person measures. Therefore, from the comparison of pre- to post-rehabilitation item measure estimates in Figure 2, the lack of effect of DIF on mean square fit statistics, and the significant relative DIF limited to a small number of items in the racked data analysis, we conclude that for the VA LV VFQ-48, DIF does not play a significant role in the outcome measures and the effects of rehabilitation are manifest as person measure changes.

Based on this conclusion, we chose to stack the pre- and post-rehabilitation data to analyze the effects of rehabilitation on person measures of visual ability. By stacking the data, we constrained the item measures to be the same before and after rehabilitation and forced all effects of rehabilitation to manifest in the person measures. Figure 4 illustrates that the item measures from the analysis of the stacked data agreed with both pre- and post-rehabilitation item measures.

Similarly, Figures 5a and 5b illustrate that the estimates of person measures from the stacked data agree with the person measure estimates for pre- and post-rehabilitation data analyzed separately.

Evaluation of the Effects of Vision Rehabilitation

Figures 6a–c compare pre- to post-rehabilitation person measure distributions for the inpatient BRC program (inpatient group; Fig. 6a), combined outpatient low-vision programs outpatient group) group (Fig. 6b), and the control group that did not receive rehabilitation services between measures (Fig. 6c).

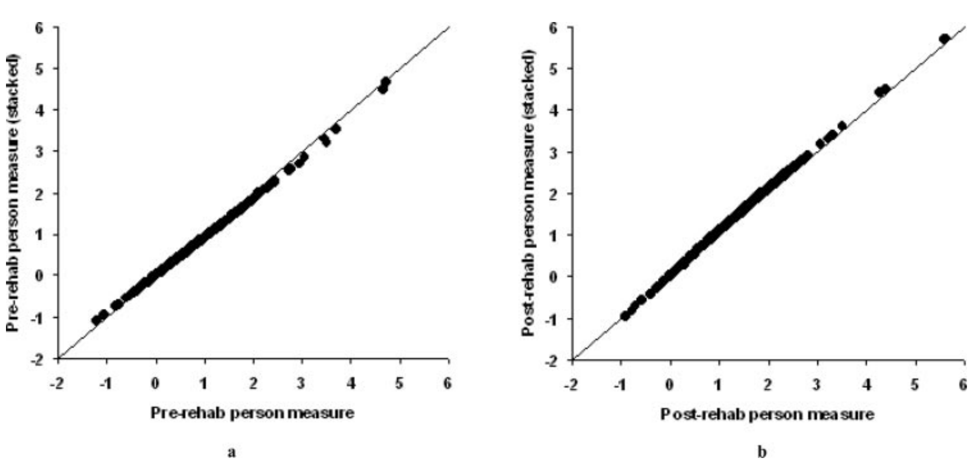

FIGURE 5. The scatterplots compare estimates of person measures calculated from the stacked data with person measures estimated for pre- and post-rehabilitation data analyzed separately. The person measures from stacked data agreed with the person measures from (**a**) pre- and (**b**) post-rehabilitation data.

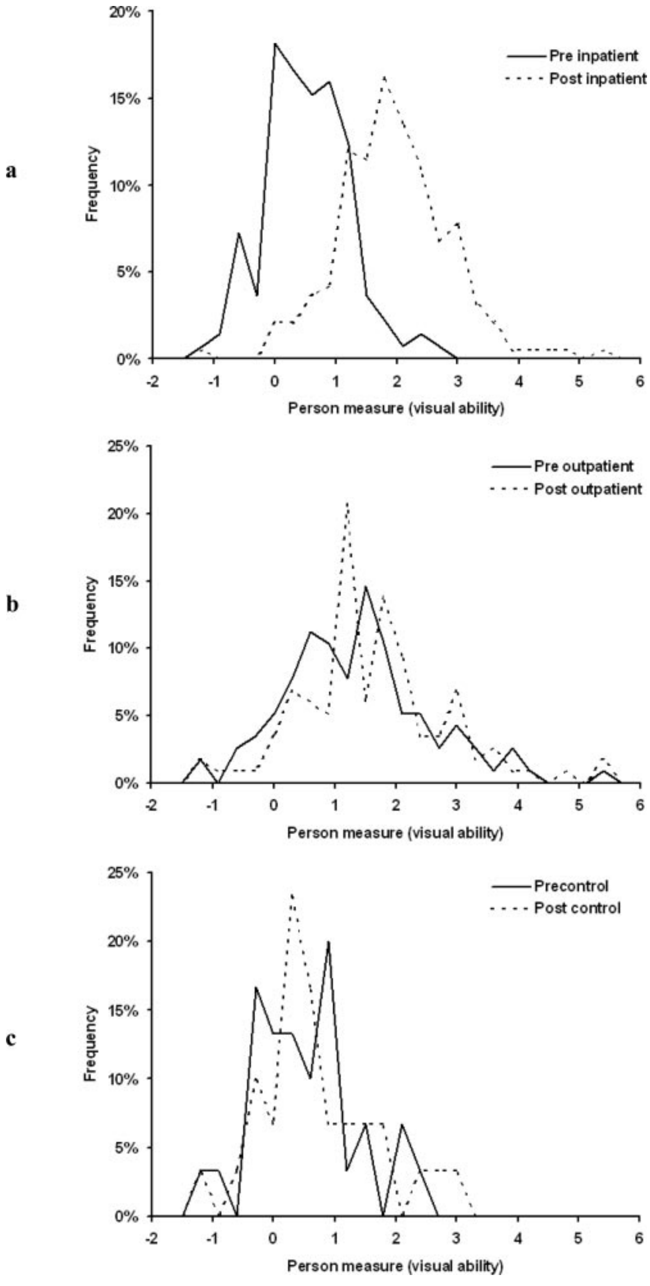

Figure 6. (**a**–**c**) Person measure distributions pre- to post-rehabilitation are compared for the inpatient BRC program, the outpatient low-vision programs combined, and the test–retest group. (**a**) In the inpatient BRC program, the average post-rehabilitation person measure (visual ability = 2.18 logits) is significantly greater than the average pre-rehabilitation person measure (visual ability = 0.66 logit; two-tailed paired t-test $P = 4.1 \times 10^{-48}$). (**b**) In the outpatient low-vision programs combined, the average post-rehabilitation person measure (visual ability = 1.85 logits) is significantly greater than the pre-rehabilitation person measure (visual ability = 1.52 logits; two-tailed paired t-test $P = 0.0022$). In the test–retest group, the average post-rehabilitation person measure (visual ability = 0.87 logit) is not significantly different from the average pre-rehabilitation person measure (visual ability = 0.71 logit; two-tailed paired t-test $P = 0.064$).

Figure 7a–c are scatterplots of pre- versus post-rehabilitation person measures for the inpatient, outpatient, and control groups, respectively.

For the inpatient group, the average person measure is 0.66 logit (SD = 0.69) before rehabilitation and 2.18 logits (SD = 0.76 logit) after rehabilitation (two-tailed paired t-test $P < 0.0001$). For the outpatient group, the average person measure was 1.52 logit (SD = 1.18) before rehabilitation and 1.85 logit (SD = 1.34 logit) after rehabilitation (two-tailed paired t-test $P = 0.0022$). For the pre-rehabilitation control group, the average person measures were 0.71 logit (SD = 0.86) for the initial test and 0.87 logit (SD = 0.93) for the retest (2-tailed paired t-test $P = 0.064$).

In the inpatient group (Fig. 7a), 10 patients (7%) fell on the identity line (no change in visual ability post-rehabilitation), but the other 129 patients (93%) fell significantly above the identity line (improved visual ability after rehabilitation). Post-rehabilitation person measures are significantly correlated with pre-rehabilitation person measures for the inpatient group ($r = 0.42$; $P < 0.0001$). Similarly, in the outpatient group (Fig. 7b), 43 patients (37%) fell on the identity line (no change in visual ability after rehabilitation), 14 patients (12%) fell significantly below the identity line (decreased visual ability after rehabilitation), and the remaining 59 patients (51%) fell significantly above the identity line (improved visual ability after rehabilitation). For the outpatient group, post-rehabilitation person measures correlated significantly with pre-rehabilitation person measures ($r = 0.61$; $P < 0.0001$). As reported previously,[9] the control group exhibited a strong correlation between test and retest person measures ($r = 0.86$; $P < 0.0001$) with no significant change in visual ability between measurements.

The change in average person measure after rehabilitation is 1.49 logits for the inpatient group and 0.33 logit for the outpatient group. These changes represent effect sizes[31] of 1.9 (97th percentile) for the inpatient group and 0.29 (62nd percentile) for the outpatient group (corrected for paired comparisons). If accepted at face value, this nearly sevenfold difference in effect size could lead us to the conclusion that the inpatient program is far more effective than the outpatient programs. However, the average pre-rehabilitation visual ability is 0.66 logit for the inpatient group and 1.52 logit for the outpatient group (a difference of 0.86 logit), whereas the average post-rehabilitation visual ability is 2.18 logit for the inpatient group and 1.85 logit for the outpatient group (a difference of 0.33 logit). The pre-rehabilitation differences in average visual ability most likely reflect differences in visual impairment eligibility criteria for the two programs: Patients must be legally blind to enter the inpatient BRC program, but the outpatient low-vision rehabilitation programs accept all patients; consequently, visual impairments range more broadly from mild to profound.

Discussion

In the case of DIF, patients interpret the item differently before and after rehabilitation. Before rehabilitation, the patient is judging the difficulty of the item based on his or her customary way of performing the activity. After rehabilitation, the patient is judging the difficulty of the activity using dispensed low-vision devices and newly acquired skills that might be item-specific. Low-vision devices in particular are targeted to change the difficulty of performing specific tasks and would not be expected to have an effect on the difficulty of other tasks. When DIF occurs, before and after rehabilitation responses must be compared for each item separately using statistical techniques that take into consideration multiple comparisons (e.g., Bonferroni correction). The power of the statistical analysis is reduced, and larger sample sizes are required to compensate for the multiple comparisons. It is more efficient and desirable to express all the rehabilitation effect as a person measure change. The VA LV VFQ-48 does not demonstrate significant DIF when comparing pre- to post-rehabili-

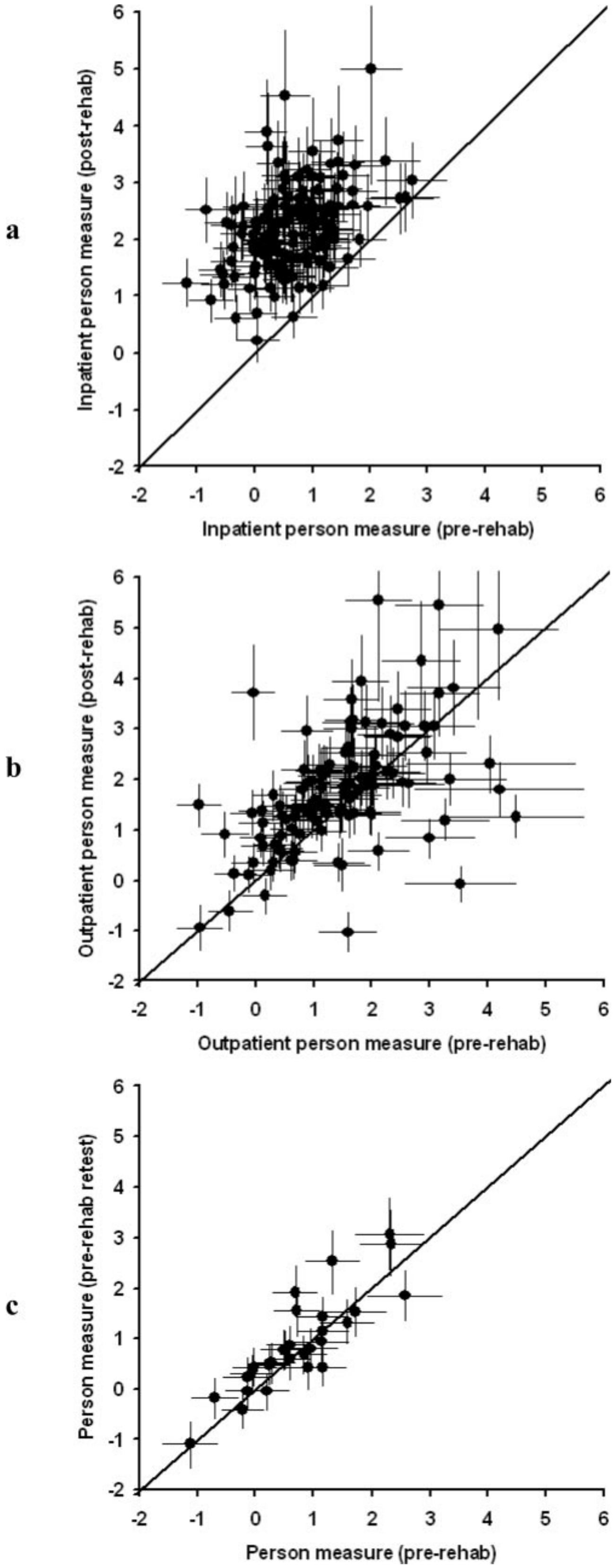

FIGURE 7. Scatterplots compare pre- and post-rehabilitation person measures for the inpatient BRC program, outpatient low-vision programs combined, and the test–retest group. (**a**) In the inpatient BRC, 7% fell on the identity line (no change in visual ability) and 93% fell above (improved visual ability post-rehabilitation). (**b**) For the outpatient low-vision programs combined, 51% fell above the identity line (improved visual ability post-rehabilitation), 37% fell on it (no change in visual ability post-rehabilitation), and 12% fell significantly below it, indicating decreased visual ability post-rehabilitation. (**c**) The test–retest group demonstrated no significant change in visual ability between measurements.

tation measures. Therefore, we can use the VA LV VFQ-48 to derive a single number, the person measure. Person measures estimated from patient responses before and after rehabilitation can be compared, to draw conclusions about the outcome of intervention.

If a questionnaire includes numerous items that do not change with high doses of the intervention being evaluated, then the measured effect will be small, because the items that do not change will dilute the effect of the items that do. Larger sample sizes are necessary for hypothesis-testing when outcomes are small. In the present study, all the items in the VA LV VFQ-48 were responsive to vision rehabilitation, which maximizes the sensitivity of the instrument as an outcome measure. The very small amount of DIF makes it possible to use a single number, the person measure, as the primary outcome variable.

There appears to be a significant dose–response relationship when outcomes of vision rehabilitation are measured with the VA LV VFQ-48. The magnitude of the effect depends on the rehabilitation program and the level of visual ability of the patient before rehabilitation. The person measures from the outpatient low-vision rehabilitation programs show a range of effects varying from significant improvement to a significant decrease in visual ability, whereas the more intense inpatient BRC program with the more visually impaired patients shows improvement for almost all patients.

Besides the obvious difference in program intensity, there are important differences between the patients served by the inpatient BRC and patients served by the outpatient low-vision rehabilitation programs. The mean pre-rehabilitation visual ability of patients in the outpatient low-vision programs is significantly higher (1.52 logits) than the pre-rehabilitation visual ability of patients in the inpatient BRC (0.66 logits). Figure 7b demonstrates that 12% of the patients in the outpatient low-vision programs decreased in visual ability after rehabilitation. It is likely that the decreases in visual ability are due to disease progression. Future studies with visual acuity measured concurrently with outcomes are needed to verify decreases in visual acuity that may occur during outcomes studies.

Comparison of outcomes from the inpatient BRC and the outpatient low-vision rehabilitation programs suggests that rehabilitation ceiling effects exist. The magnitude of change in person measure that can be achieved through vision rehabilitation may not depend solely on the dose of rehabilitation or the types of services and assistive devices that are provided. The distribution of pre-rehabilitation person measures may also limit the magnitude of change. Patients with more visual ability (less difficulty performing daily living tasks) before rehabilitation may have lower rehabilitation demand (less need for rehabilitation) than patients with lower visual ability (more difficulty performing daily activities) before rehabilitation. One consequence of a rehabilitation ceiling effect is that when low-vision outcome studies are conducted with patients having higher visual ability (less difficulty performing tasks) larger sample sizes will be necessary for hypothesis testing.

In conclusion, we have shown that in addition to being a valid and reliable measure of visual ability,[9] the VA LV VFQ-48 is a sensitive measure of changes that occur in visual ability as a result of rehabilitation. Patients' self reports of the difficulty they experience performing daily activities measured with this instrument can be used to compute a single number, the person measure that can serve as an outcome measure in clinical studies. The change in person measure from before to after rehabilitation is an expression of vision rehabilitation outcomes, because there is very little DIF from before to after rehabilitation. The VA LV VFQ-48 is responsive to the effects of inpatient VA blind rehabilitation and outpatient low-vision rehabilitation programs in both the VA and private sector. The VA LV VFQ-48 can be used to compare programs that offer different levels of intervention and serve patients across the continuum of vision loss.

References

1. McCabe P, Nason F, Demers-Turco P, Friedman D, Seddon JM. Evaluating the effectiveness of a vision rehabilitation intervention using an objective and subjective measure of functional performance. *Ophthalmic Epidemiol*. 2000;7:259–270.
2. Wolffsohn JS, Cochrane AL. Design of the Low Vision Quality-of-Life Questionnaire (LVQOL) and measuring the outcome of low vision rehabilitation. *Am J Ophthalmol*. 2000;130:793–802.
3. Reeves BC, Harper RA, Russell WB. Enhanced low vision rehabilitation for people with age related macular degeneration: a randomized controlled trial. *Br J Ophthalmol*. 2004;88:1443–1449.
4. Scott IU, Smiddy WE, Schiffman J, Feuer WJ, Pappas CJ. Quality of life of low vision patients and the impact of low vision services. *Am J Ophthalmol*. 1999;128:54–62.
5. Brody BL, Roch-Levecq AC, Gamst AC, Maclean K, Kaplan RM, Brown SI. Self-management of age related macular degeneration and quality of life. *Arch Ophthalmol*. 2002;120:1477–1483.
6. Scilley K, DeCarlo DK, Wells J, Owsley C. Vision-specific health-related quality of life in age related maculopathy patients presenting for low vision services. *Ophthalmic Epidemiol*. 2004;11:131–146.
7. Massof RW. Likert and Guttman scaling of visual function rating scale questionnaires. *Ophthalmic Epidemiol*. 2004;11:381–399.
8. Stelmack JA, Stelmack TR, Massof RW. Measuring low vision rehabilitation outcomes with the NEI VFQ-25. *Invest Ophthalmol Vis Sci*. 2002;43:2859–2868.
9. Stelmack JA, Szlyk JP, Stelmack TR, et al. Psychometric properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire. *Invest Ophthalmol Vis Sci.* 2004;35:3919–3928.
10. Stelmack J, Szlyk J, Stelmack T, et al. Use of Rasch person item map in exploratory data analysis: a clinical perspective. *J Rehabil Res Dev*. 2004;41:233–242.
11. 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.
12. Stelmack J, Rosenbloom A, Brenneman C, Stelmack T. Patient's perceptions of the need for low vision devices. *J Vis Impair Blind*. 2003;97:521–535.
13. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum data sets. *Med Care.* 2004;42(suppl):1–24.
14. Velozo CA, Lai KS, 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;4:6667–6680.
15. 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.
16. Pesudovs K, Garamendi E, Keeves JP, Elliott DB. The activities of daily vision scale for cataract surgery outcomes: reevaluating validity with Rasch analysis. *Invest Ophthalmol Vis Sci*. 2003;44:2893–2899.
17. Turano KA, Massof RW, Quigley HA. A self-assessment instrument designed for measuring independent mobility in retinitis pigmentosa patients: generalizability to glaucoma patients. *Invest Ophthalmol Vis Sci*. 2002;43:2874–2881.
18. 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.
19. 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.
20. Massof RW, Fletcher DC. Evaluation of the NEI visual functioning questionnaire as an interval measure of visual ability in low vision. *Vis Res*. 2001;41:397–413.
21. Bond TG, Fox CM. *Applying the Rasch Model.* Mahwah, NJ: Earlbaum; 2001.
22. Massof RW. The measurement of vision disability. *Optom Vis Sci*. 2002;79:515–552.
23. Massof RW. A systems model for low vision rehabilitation I. Basic concepts. *Optom Vis Sci*. 1995;72:725–736.
24. Massof RW. A systems model for low vision rehabilitation II. Measurement of vision disabilities. *Optom Vis Sci*. 1998;75:349–373.
25. Wright BD. Rack and stack: time 1 vs. time 2. *Transactions of the Rasch Measurement SIG American Educational Research Association*. 2003;17:905–906.
26. Forster PA, Dawson VM, Reid D. Measuring preparedness to teach with ICT. *Australasian J Educ Tech*. 2005;21:1–18.
27. Dunn CM, Chadwick G. Code of Federal Regulations. In: Allen W, ed. *Protecting Study Volunteers in Research.* Boston, MA: Center Watch, Inc.; 1999:203.
28. Andrich D. A rating formulation for ordered response categories. *Psychometrika.* 1978;43;561–573.
29. Linacre JM, Wright B. *A User's Guide to WINSTEPS Rasch-Model Computer Program.* Chicago, IL: MESA Press; 2001.
30. Bond TG. Validity and assessment: a Rasch measurement perspective. *Metodol Ciencias Comport*. 2003;5:179–194.
31. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Erlbaum; 1988.