

# The Impact of Vision Impairment Questionnaire: An Evaluation of Its Measurement Properties using Rasch Analysis

Ecosse L. Lamoureux,<sup>1</sup> Julie F. Pallant,<sup>2</sup> Konrad Pesudovs,<sup>3</sup> Jennifer B. Hassell,<sup>1</sup> and Jill E. Keefe<sup>1,4</sup>

**PURPOSE.** To explore the psychometric properties of the Impact of Vision Impairment scale (IVI) by using Rasch analysis.

**METHODS.** Three hundred fourteen first-time referrals to low-vision clinics completed the 32-item IVI. The data were Rasch-analyzed with a partial credit model using RUMM2020 software (RUMM Laboratory, Perth, WA, Australia). The overall fit of the model, response scale, individual item fit, differential item functioning, unidimensionality, and person-separation reliability were assessed.

**RESULTS.** Initially, 26 items displayed disordered thresholds. However, collapsing the response scale to three categories (4 items) and four categories (28 items) produced ordered response thresholds for all items. Four items with high proportions of missing responses, poor spread, high skewness, and deviation between observed and expected model curves were then removed. This adjustment produced overall fit to the Rasch model (item-trait interaction  $\chi^2 = 118.3$ ;  $P = 0.32$ ). The final mean (SD) person and item fit residuals are 0.06 (0.85) and  $-0.20$  (1.45), respectively. The person-separation reliability was 0.9, indicating that the scale was able to discriminate between several different groups of participants. The revised scale was well targeted to the participants, with similar mean locations for items (0.00) and persons (0.16). A significant difference between participants of mild, moderate, and severe visual impairment (ANOVA;  $P < 0.001$ ) supported the criterion validity of the Rasch-scaled IVI.

**CONCLUSIONS.** The results provide support for the measurement properties of the Rasch-scaled 28-item version of the IVI and of its potential for assessing outcomes of low-vision rehabilitation. A raw score-to-Rasch person measure conversion is supplied. (*Invest Ophthalmol Vis Sci.* 2006;47:4732-4741) DOI: 10.1167/iovs.06-0220

---

From the <sup>1</sup>Centre for Eye Research Australia, The University of Melbourne, Melbourne, Victoria, Australia; the <sup>2</sup>Swinburne University of Technology, Melbourne, Victoria, Australia.; the <sup>3</sup>National Health and Medical Research Council Centre for Clinical Eye Research, Flinders University and Flinders Medical Centre, Adelaide, South Australia, Australia; and <sup>4</sup>Vision Cooperative Research Center, Sydney, New South Wales, Australia.

EL is a current National Health and Medical Research Council (Canberra, Australia) Public Health Fellow (NHMRC Grant 359277).

Submitted for publication February 28, 2006; revised June 11 and 23, 2006; accepted September 19, 2006.

Disclosure: E.L. Lamoureux, None; J.F. Pallant, None; K. Pesudovs, None; J.B. Hassell, None; J.E. Keefe, 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: Ecosse L. Lamoureux, Centre for Eye Research Australia, Department of Ophthalmology, University of Melbourne, 32 Gisborne Street, East Melbourne, Victoria, 3002, Australia; ecosse@unimelb.edu.au.

Low vision has been defined as any chronic visual impairment, not correctable by ordinary spectacles or contact lenses that impacts on daily living.<sup>1</sup> To provide appropriate rehabilitation programs to individuals with low vision, it is necessary to assess their difficulty in participating in daily living. Several vision-specific questionnaires have been developed to serve as patients' self-evaluation tools, and the psychometric characteristics of most of these instruments have been reviewed.<sup>2-4</sup> One of them, the 32-item Impact of Vision Impairment (IVI) questionnaire,<sup>5</sup> has been used to assess the determinants of participation in daily living for individuals with low vision,<sup>6</sup> to evaluate the impact of diabetic retinopathy and age-related macular degeneration on participation in daily activities,<sup>7,8</sup> and to determine the association between glaucomatous visual field loss and activities of daily living.<sup>9</sup>

The IVI questionnaire provides six response categories for each item (ranging from *not at all* to *can't do because of eye sight*) and employs Likert scoring. Although it is implied that Likert values are monotonic with the latent trait they are endeavoring to assess, it is difficult to confirm that they possess an interval measurement component. The validity of the Likert scale, as representing an interval scale, has been questioned by proponents of Item Response Theory—in particular, the application of Rasch analysis.<sup>10-14</sup> Rasch analysis offers an elegant approach to addressing several important methodological characteristics associated with scale development and construct validation, as well as providing a transformation of the ordinal raw scores to a linear interval scale permitting the use of parametric statistical techniques.<sup>15</sup> Rasch analysis also calculates item difficulty in relation to person difficulty and assesses the scale validity—in particular, the item and person fit to the overall construct.<sup>16</sup>

Although there are currently several visual functioning questionnaires available, only a few have been developed using Rasch analysis.<sup>17-24</sup> Others have been Rasch assessed, allowing for improvements to be made to their structure.<sup>16,25-27</sup> With the exception of scales such as the Veterans Affairs Low Vision Visual Functioning Questionnaire (LV VFQ-48),<sup>18</sup> few questionnaires have the range of items to assess difficulty with daily activities and a demonstrated capacity to evaluate activities subsequent to low-vision rehabilitation. The IVI has been designed to assess the restriction of participation in daily living as well as the effectiveness of rehabilitation outcomes in low vision, unlike most vision-specific questionnaires that typically assess visual functioning. To determine whether the IVI possesses the measurement characteristics (interval scale, validity, and reliability) and is an accurate and sensitive evaluation instrument for vision rehabilitation, we used Rasch analysis on the IVI, considering item reduction if necessary.

## METHODS

The data for this study were from first-time referrals to low-vision rehabilitation centers across Victoria (Australia). An ophthalmologist's report, providing the cause of vision loss and visual acuities, was

required for each participant. The eligibility criteria for the study included best presenting visual acuity  $<6/12$  (or  $6/12$  or better with restricted fields),  $\geq 18$  years of age, and the ability to converse in English. Participants signed a consent form, and low-vision rehabilitation files were accessed to obtain clinical data. Ethical approval was obtained from the Royal Victorian Eye and Ear Hospitals Human Research and Ethics Committee. This research adhered to the tenets of the Declaration of Helsinki. Sociodemographic and clinical data were collected.

## IVI Questionnaire

A detailed description of the IVI questionnaire has been fully published elsewhere<sup>5</sup> and is summarized herein. The questionnaire was developed in three stages. Initially, focus groups, comprising individuals with the most common causes of impaired vision, identified activities causing restriction of participation to daily living.<sup>28</sup> In the second stage, issues identified in focus groups were operationalized into a bank of 76 items. Existing instruments—namely, the Activities of Daily Vision Scale,<sup>29</sup> the Visual Function Questionnaire,<sup>30</sup> the National Eye Institute-Visual Function Questionnaire (NEI-VFQ),<sup>31</sup> and the Bristol Vision-Related Quality of Life (VQOL)<sup>32</sup>—were reviewed for content and scaling relevant to the issues identified in the focus groups. Items pertinent to ocular symptoms and a person's limitation in activities—for example, in seeing small objects—were not included in the IVI. In the third stage, the IVI was trialed in two consecutive versions before the final 32-item version was derived. All versions of the IVI retained the core 10 questions of the VQOL, and all retained questions in each of the five provisional domains: mobility; household; personal care; consumer and social interactions; and leisure and work and emotional reaction to vision loss.

In this study, the 32-item IVI instrument was either self- or interviewer-administered, as high levels of consistency have been recorded in both methods.<sup>5</sup> Proxy answers were not solicited from caregivers or relatives, to avoid biasing the IVI responses to the perception of another person's opinion of the participant's ability. Responses to the IVI items were rated with a five-category Likert scale: *not at all* (0), *hardly at all* (1), *a little* (2), *a fair amount* (3), *a lot* (4), and *can't do because of eyesight* (5), with an additional response category, *don't do because of other reasons*, for 19 items. The later response was not included in computing the average overall or domain score. The wording preceding these items was: *In the past month, how much has your eyesight interfered with the following activities*. For the remaining 13 items, the rating scale used was: *not at all* (0), *very rarely* (1), *a little of the time* (2), *a fair amount of the time* (3), *a lot of the time* (4), and *all the time* (5). The wording preceding these items was: *In the past month, how often has your eyesight made you concerned or worried about the following*. Data for this study represent information collected at baseline as part of an intervention study between 2001 and 2002.

## Rasch Analysis

The Rasch model was named after the Danish Mathematician Georg Rasch.<sup>33</sup> The model specifies what should be an expected pattern of responses to items if measurement (at the interval level) is to be achieved. For the Rasch model, dichotomous<sup>33,34</sup> and polytomous<sup>34</sup> versions are available. The response patterns achieved are tested against what is expected; a probabilistic form of Guttman scaling<sup>35</sup> and a variety of statistics are used to assess fit to the model.<sup>36</sup>

The Rasch model assumes that the probability of a given respondent's affirming an item is a logistic function of the relative distance between the item's location and the respondent's location on a linear scale. If a person's ability in performing a particular activity is lower than the required ability for that particular task, the probability of the person's rating the task in the highest scoring category (in this case: *can't do because of eyesight*) is high. Conversely, if a person's level of ability is greater than the ability required for a particular task, the probability of the person's rating the task in the low scoring category

(e.g., *not at all*) is high. Hence, it is expected that the probability of using any particular rating category will increase monotonically with the difference between the person's level of difficulty in performing daily activities and the level of difficulty required for the particular task.

For ease of interpretation of scores the IVI rating scale scoring was reversed for the Rasch analysis (0 as 5, 1 as 4, 2 as 3, 3 as 2, 4 as 1, and 5 as 0). A positive item, measured in logits (the unit of measure used by Rasch for calibrating items and measuring persons) on the Rasch scale indicates that the item requires a higher level of participation than the mean of the items, whereas a negative item logit suggests that the item requires a lower level of participation than the average. A positive person-logit score suggests that the person's level of participation is higher than the mean required level of difficulty for the items. Conversely, if a person-logit score is negative, the person's perceived level of participation is lower than the average required level of difficulty.

The data were evaluated for fit to the Rasch model using the RUMM2020 (Rasch unidimensional measurement models; RUMM Laboratory, Perth, WA, Australia) software,<sup>37</sup> with the goal of assessing how well the observed data fit the expectations of the measurement model. The partial-credit approach<sup>38</sup> (which allows each item to have its own threshold parameters) was used because the likelihood-ratio test was statistically significant ( $P < 0.001$ ) indicating that the rating scale model (which requires equivalent thresholds across all items) was not appropriate. The likelihood-ratio test was still statistically significant ( $P < 0.001$ ) when applied to the two subsets of items (19 and 13 items), suggesting that the partial-credit approach was more suitable. Three overall fit statistics were considered. Two were fit residuals statistics, which represent the residuals between the expected estimate and actual values for each person-item, summed over all items for each person and over all persons for each item. The residuals are transformed to approximate a z-score and represent a standardized normal distribution where perfect fit to the model would have a mean of approximately 0 and an SD of 1. An item-trait interaction score reported as a  $\chi^2$ , which reflects the property of invariance across the trait, was also provided. A statistically nonsignificant probability value ( $P > 0.05$ ) indicates no substantial deviation from the model. Individual item or person statistics where fit residuals values  $>2.5$  or probability values below the Bonferroni adjusted  $\alpha$  value (i.e.,  $0.05/32 = 0.001$ ) are also used to indicate misfitting to the model. In addition to these overall fit statistics, the RUMM2020 program provides an indication of person-separation reliability using the person-separation index (PSI range, 0–1) which indicates how well the items of the instrument separate, or spread out, the subjects in the sample. A person-separation reliability value from RUMM of 0.7 is the equivalent of a G value of 1.5, representing the ability to distinguish two distinct strata of person ability.<sup>39,40</sup> A value of 0.9 is equivalent to a G value of 3, with the ability to distinguish four strata of person ability.

Misfit of items indicates a lack of the expected probabilistic relationship between the item and other items in the scale. This introduces noise into the measurement, diminishing the instrument's quality. In the event of item misfit, two strategies were undertaken to improve the scale. First, a lack of ordered responses (disordered thresholds) was determined. Disordered thresholds occur when the response is selected by participants, with a wide range of abilities on the underlying trait being measured, or a person location between category boundaries will not give that category the greatest probability of being observed. This can occur when there are too many response options, or when the labeling of options is similar to one another, potentially confusing or open to misinterpretation (e.g., *not at all*, *hardly at all*, and *a little*). Collapsing the categories where disordered thresholds occur can often improve overall fit to the model. Initially, the RUMM2020 software will identify disordered thresholds. Thereafter decisions will be made how best to collapse categories (e.g., rescore a five-point scale into a four-point scale). A visual examination of the way in which categories are working will indicate possible ways to collapse categories. For example, if a category is less likely to be chosen or not appropriately used across the whole scale, it could be collapsed with

TABLE 1. Characteristics of the 314 Study Participants

Age (y)	
Mean $\pm$ SD	78.4 $\pm$ 12.9
Range	21-102
Gender	
Men	114 (37%)
Women	200 (63%)
Presenting visual acuity	
<6/12-6/18	131 (42%)
<6/18-6/60	147 (47%)
<6/60	36 (11%)
Near vision	
N8 or better	151 (50%)
<N8-N20	92 (31%)
<N20-N48	35 (12%)
<N48	22 (7%)
Main cause of vision loss	
Age-related macular degeneration	169 (54%)
Diabetic retinopathy	49 (16%)
Glaucoma	38 (12%)
Other	58 (18%)
Median (min, max)	3 (0,84)
Comorbidity	
Yes	258 (82%)
No	56 (18%)
Not at all	61 (24%)
A little	91 (35%)
A great deal	106 (41%)

All data except those for age are number of subjects (%).

the adjacent category. Where alternative collapsing strategies seem possible, that pattern which produces the best fit for the item is chosen.

Once disordered thresholds are removed, fit of data to the Rasch model is assessed by examining deviations from model expectations, including DIF (differential item functioning). DIF occurs when different groups or person factor within the sample (e.g., degree of visual impairment)—despite equal levels of the underlying characteristic being measured (participation in daily living)—responds in a different manner to an individual item. DIF can be detected both graphically, by inspection of the item characteristic curves, and statistically, by using analysis of variance comparing scores across each level of the person factor and across different levels of trait (referred to as class intervals).

Once fit of the data to the Rasch model is determined by the appropriate range of statistics at the model, individual item and person levels as described earlier, it is necessary to confirm that the scale is

appropriately targeted to the population being assessed. It is also important to confirm the unidimensionality of the questionnaire using principal components analysis (PCA) of the residuals available in RUMM. Unidimensionality is important because it provides further evidence that the instrument is measuring the underlying trait that it is believed to measure. This is demonstrated when there are no associations in the residuals derived from the difference between observed values and model expectations (local independence). Unidimensionality is formally tested by allowing the pattern of factor loadings on the first residual to determine subsets of items. If person estimates derived from these subsets of items differ significantly (using the *t*-test) from the estimates derived from the full scale, a breach of the assumption of local independence is indicated.<sup>41</sup>

## RESULTS

### Participants' Characteristics

The mean age of the 314 participants was 78.4  $\pm$  12.9 years (Table 1). The majority (63%) were women and had age-related macular degeneration (54%).

### Overall Fit

To maximize the retention of the initial character of the IVI, a minimalist approach was taken to item and scale changes such that only those changes necessary to improve scale functioning were made. The initial fit of the data to the Rasch model showed a significant item-trait interaction ( $\chi^2 = 246$ ,  $P < 0.001$ ), suggesting misfit between the data and model. The mean (SD) fit residual values were 0.42 (1.16) for items and  $-0.27$  (1.68) for persons. Ideally, the mean and SD are expected to be closer to 0 and 1, respectively, suggesting misfit to the model by items and respondents. The person-separation reliability was 0.95.

The pattern of item thresholds was first examined for disordering, suggesting that the participants could not reliably discriminate between the categories of difficulty. Disordered threshold is a violation of the measurement construct, in that there is discordance between the category probabilities and the underlying trait. Twenty-six items were found with disordered thresholds. An examination of items with disordered thresholds indicated that not all response categories had a point along the ability continuum where they were the most likely response. For example, for item mob22 *safety outside of home* (Fig. 1), response categories 1 and 4 do not have a range

mob22 Safety outside of home    Locn = -0.208    Unit = 0.207    FitRes = 0.049    ChiSq(Pr) = 0.014    SampleN = 314

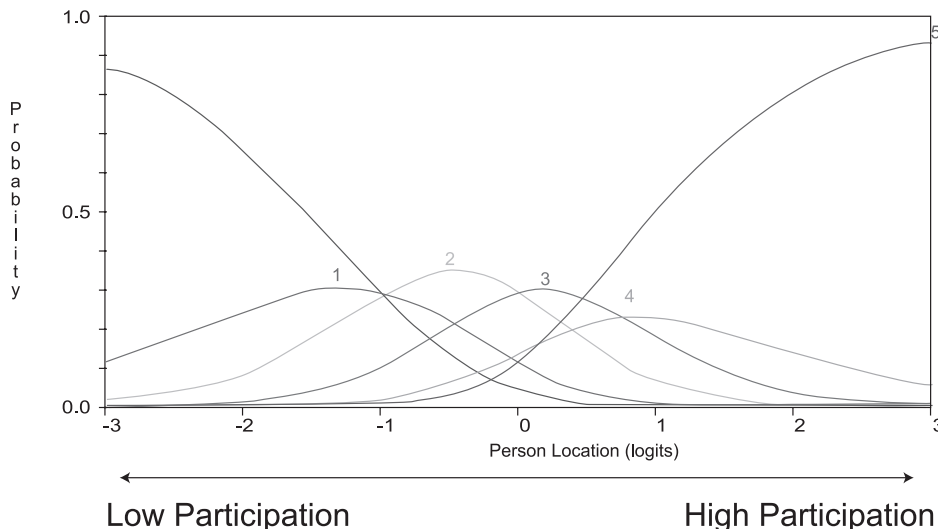


FIGURE 1. Category probability curve showing a disordered threshold for item safety outside the home.

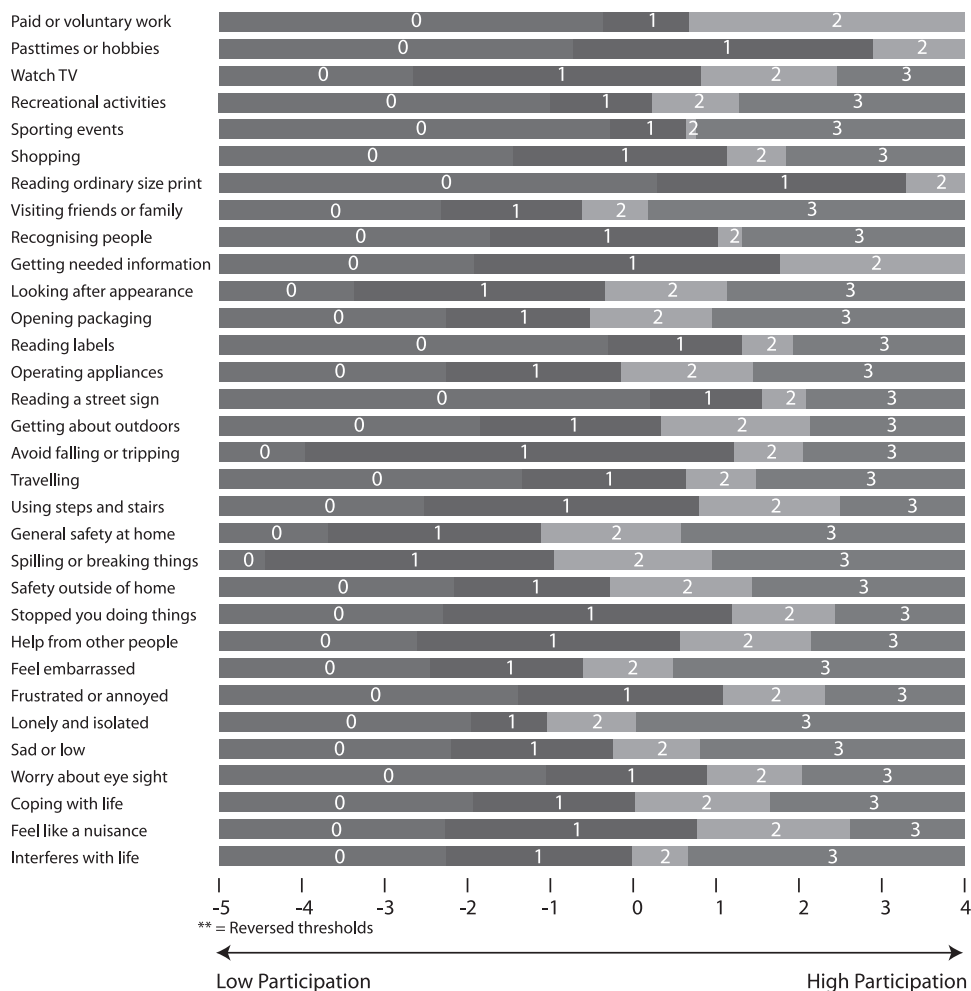


FIGURE 2. Threshold map of the IVI questionnaire, showing ordered thresholds after item rescoring.

along the ability scale where they are the category most likely to be chosen.

Consequently, scores for the 32 items were recoded by collapsing six categories to four categories (coded 3, 2, 2, 1, 1, 0). This resulted in an improvement in the overall model fit, as indicated by a change in the item-trait probability values. Following rescoring, only four items had disordered thresholds. Inspection of these showed overlapping between the second and third response categories (scores of 1 and 2) which were then collapsed, forming three categories (2, 1, 1, 1, 1, 0). This resulted in a further improvement in the overall item-trait interaction statistics, with no item showing disordered thresholds (Fig. 2).

**Estimates of Person and Item Measures**

After the recoding of the IVI items, the fit of the individual items suggested no serious misfit to model expectation (mean 0.09, SD 0.9). All items showed fit residual values in the range from -1.7 to 2.2, and no items exceeding the Bonferroni adjusted  $\alpha > 0.001$ , indicating no significant deviation from the model (Table 2). Individual person-fit statistics showed that 12 (3.8%) participants had fit residuals outside the acceptable range ( $>2.5$ ). Further analysis of the misfitting participants showed inconsistent patterns in the items where extreme responses were observed. On removal of these persons, the item-trait interaction statistics improved further ( $\chi^2 = 179; P = 0.002$ ). The person's fit residual also improved for mean (-0.27 to -0.20) and SD (1.68 to 1.47).

**Differential Item Functioning**

Within the framework of Rasch measurement, a scale should function consistently, irrespective of subgroups within the sample being assessed.<sup>42</sup> We were interested to know whether different subgroups in our sample (gender, degree of visual impairment, comorbidity, and effect of comorbidity on daily living) responded in the same way to the IVI items. We selected these subgroups as there was a substantially large proportion of women (64%) in our sample, and our previous work has shown that the other variables are related to restriction of participation in daily living.<sup>6,7,43</sup> This finding was explored in RUMM by using DIF with a Bonferroni adjusted  $P = 0.001$  (0.05/32). All items were found to be free from DIF, with probabilities exceeding the adjusted  $\alpha$  for each of the person factors assessed.

**Overall Item-Trait Interaction**

Despite effective rescoring and person and item fit residual mean and SD scores approximating 0 and 1, respectively, the item-trait interaction total value remained statistically significant ( $\chi^2 = 179; P = 0.002$ ), suggesting some remaining misfit to the model. Further removal of persons did not improve the total  $\chi^2$  and probability values. A minimalist approach for item removal was therefore considered based on several additional criteria, including a high level of the irrelevant response category (i.e., *don't do because of other reasons*), ceiling effect (the percentage in the least-able end of the response sale), and skewness.<sup>16</sup> The item-trait interaction was used to assess scale

TABLE 2. Fit of the 32 Items to the Rasch Model after Rescoring

Item	Location	Fit Residuals	DF	$\chi^2$	Probability Score
Paid work	0.14	1.921	130	6.204	0.184
Pastimes and hobbies	1.07	0.000	285	8.994	0.061
Ability to enjoy TV	0.21	0.786	295	7.269	0.122
Taking part in recreational activities	0.15	-0.961	213	7.561	0.109
Shopping	0.35	0.246	173	7.226	0.124
Reading ordinary size print	0.5	-0.188	269	1.742	0.783
Visiting friends or family	1.77	0.186	299	1.229	0.873
Recognizing people	-0.93	0.208	274	3.165	0.531
Getting information	0.2	-0.507	298	1.334	0.856
Looking after appearance	-0.08	-0.020	294	0.785	0.940
Opening packaging	-0.86	0.313	296	3.148	0.533
Reading labels or instructions	-0.61	-0.599	291	6.075	0.194
Operating household appliances	0.97	0.510	292	10.294	0.036
Shopping	-0.33	-0.649	292	1.947	0.746
Reading a street sign	1.27	1.260	286	11.474	0.022
Getting outdoors	0.19	-1.691	282	12.657	0.013
Avoid falling or tripping	-0.25	0.282	291	7.915	0.095
Travelling or using transport	0.24	-0.909	229	7.993	0.092
Going down steps, stairs, or curbs	0.23	-0.096	285	11.708	0.020
General safety at home	-1.42	0.161	298	1.885	0.757
Spilling or breaking things	-1.48	0.506	294	5.165	0.271
General safety outside of home	-0.34	-0.437	289	10.214	0.037
Stops from doing things	0.43	-1.244	295	6.378	0.173
Needs help from other people	0.02	-0.363	297	1.414	0.842
Embarrassed	-0.88	1.869	292	1.661	0.798
Frustrated or annoyed	0.66	-0.215	292	6.146	0.189
Lonely and isolated	-0.99	0.905	293	13.211	0.010
Sad or low	-0.55	0.575	293	6.468	0.167
Worried eyesight is getting worse	0.61	2.231	292	6.743	0.150
Coping with life	-0.1	0.335	295	2.655	0.617
Feels like a nuisance	0.35	-1.499	293	5.636	0.228
Interfers with life in general	-0.55	-0.009	294	5.539	0.236

All items showed Fit Residuals values  $<2.5$  and Bonferroni adjusted probability scores  $>0.001$ . DF = Degrees of freedom.

functioning, rather than the person-separation reliability value. In addition, all items were viewed graphically to determine how well the observed model tended to fit the expected model curve in groups of responders across the trait (called class intervals). Items with good fit tend to show each of the group plots lying on the curve. Those with plots that were steeper than the curve would be considered to be overdiscriminating and those flatter than the curve, underdiscriminating.<sup>15</sup> The items *paid or voluntary work* and *going out to sporting events* had the highest proportions of irrelevant responses (55.4% and 41.4%). The items *favorite pastimes or hobbies* and *reading a sign across the street* had ceiling effect (70%–75%).

In addition, these four items showed deviations from the model curves compared with the remaining items (see Fig. 3, showing items *paid or voluntary work* and *going out to sporting events*). Item reduction was an iterative procedure, with one item removed at a time and fit re-estimated accordingly. The item with the highest number of candidate criteria (irrelevancy, spread, skewness, and poor fit to the expected curve) ordered by priority, was removed first. Consequently, the following items were removed individually in the following order: Lei1, *paid or voluntary work*; Lei5, *going out to sports, movies, or plays*; Lei2, *favorite pastimes or hobbies*; and mob15, *reading a sign across the street*.

The item-trait interaction total statistics consistently improved after each consecutive removal but only reached a statistically nonsignificant level after all four items were removed ( $\chi^2 = 118$ ,  $P = 0.32$ ). The final mean person and item fit residual values were 0.068 (SD 0.85) and  $-0.203$  (SD 1.45), respectively. The person separation reliability score was 0.95,

which indicates that the scale is able to discriminate between several different groups of participants.

### Person–Item Map

The person-item map shown in Figure 4 displays the participants scores on the Rasch calibrated scale (on the lefthand side) and shows the relative difficulty levels of each of the IVI items on the righthand side. Participants having the highest level of participation and the most difficult items are at the top of the diagram. Conversely, the participants having the lowest level of participation and the least difficult items are at the bottom. Estimates of the participants' perceived level of participation (in logits) were not significantly different from a normal distribution (Kolmogorov-Smirnov z-test score = 0.57;  $P = 0.9$ ). There was an even spread of items across the full range of respondents' scores, suggesting effective targeting of the IVI items. In addition, the mean person location logit value (0.18) indicates that, overall, the questionnaire was well targeted, with participants on average at a marginally higher level of ability than the average of the scale items (which would be 0 logit). The five most difficult items in the revised IVI were *reading ordinary size print*, *reading labels or instructions on medicine*, *feeling frustrated or annoyed*, *worried about eyesight getting worse*, and *shopping*, with logit scores of 2.12, 1.19, 0.75, 0.74, and 0.66, respectively. Conversely, the five least difficult items were *general safety at home*, *spilling or breaking things*, *feeling lonely and isolated*, *feeling embarrassed*, and *visiting friends or family* with logit scores of  $-1.47$ ,  $-1.39$ ,  $-1.08$ ,  $-0.92$ , and  $-0.87$ , respectively.

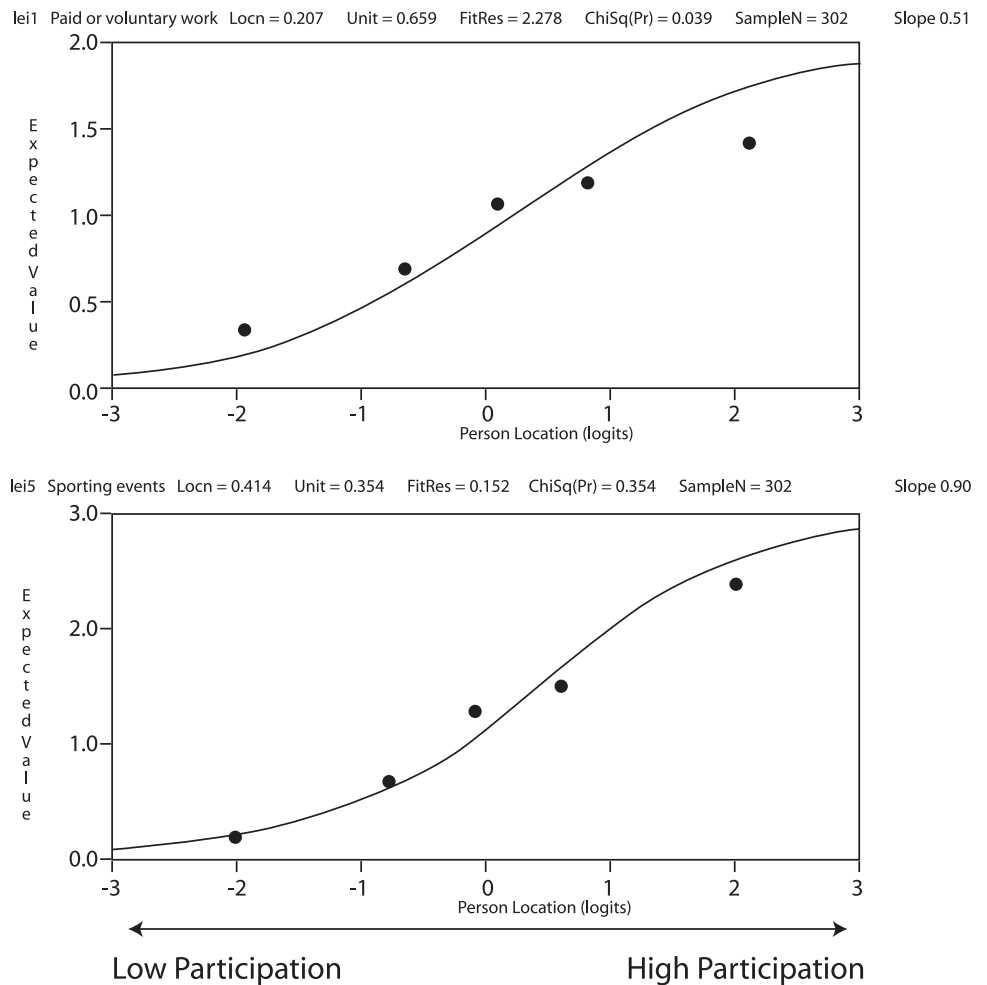


FIGURE 3. Item characteristic curves for two individual IVI questionnaire items (Lei5, *paid or voluntary work* and *going out to sporting events*) showing deviation of the observed group responses (*black dots*) from the model curves (*solid line*).

**Test of Local Independence Assumption**

A PCA of the residuals was used to assess the dimensionality of the IVI. The residuals are what remain when the Rasch factor or the underlying trait has been removed. The pattern of item loadings on the first extracted factor shows that the residuals loaded in opposite directions on two subsets defined by positive and negative loadings on the first factor (Table 3). Only those items with loading factors greater than ±0.3 were used. No significant differences were found between the person estimates of the IVI and the eight-item positive subtest (*t*-test; *P* = 0.95) and seven-item negative subset (*t*-test; *P* = 0.97). This finding suggests no breach of the assumption of local independence, therefore supporting the unidimensionality of the scale.

**Criterion Validity**

The criterion validity of the Rasch-scaled IVI was tested by assessing its ability to discriminate between participants of different levels of visual impairment—namely, mild (VA, <6/12–6/18), moderate (<6/18–6/60), and severe (<6/60). There was a significant difference between the three groups (ANOVA;  $F_{(2,265)} = 13.3$ ; *P* < 0.0001) with poorer visual acuity being associated with greater restriction of participation (1.18, 0.42, and –0.003, mean logit values for mild, moderate, and severe visual impairment, respectively).

**Scoring of the IVI Questionnaire**

Other investigators wanting to use the IVI questionnaire can use our validation data to convert raw scores into Rasch

person measures without having to perform Rasch analysis. This conversion mainly holds for patients with complete data. Raw scores are calculated by, first, reversing scores (0, 1, 2, 3, 4, 5) to (5, 4, 3, 2, 1, 0) to give the better IVI score to the less impaired, as described in the Methods section. Second, categories are collapsed to four (3, 2, 2, 1, 1, 0) or three categories (2, 1, 1, 1, 1, 0) as described earlier in the Results section. The average of the 28 items gives the IVI raw score. This score is related to the IVI Rasch person measure, as illustrated in Figure 5. The relationship is double asymptotic because the average raw rating has a floor and a ceiling (at 0 and 3). The relationship can be described by the double-asymptotic nonlinear regression<sup>44</sup>:  $IVI_{person\ measure} = 19.72 \log(IVI_{raw\ score}/3 - IVI_{raw\ score}) + 48.29$ . This equation can be used to convert raw scores to Rasch person measures.

**DISCUSSION**

Our goal was to establish whether the IVI questionnaire, which has been conventionally validated as a tool for assessing restriction of participation in daily living in visually impaired individuals, meets the formal requirements of measurement as defined by the Rasch model. The response scale was collapsed, and four misfitting items were identified and removed. The Rasch-scaled 28-item IVI demonstrates a justifiable scale for measuring perceived restriction of participation in daily activities for individuals with impaired vision. It also possesses high reliability, demonstrated validity, and effective targeting and shows no

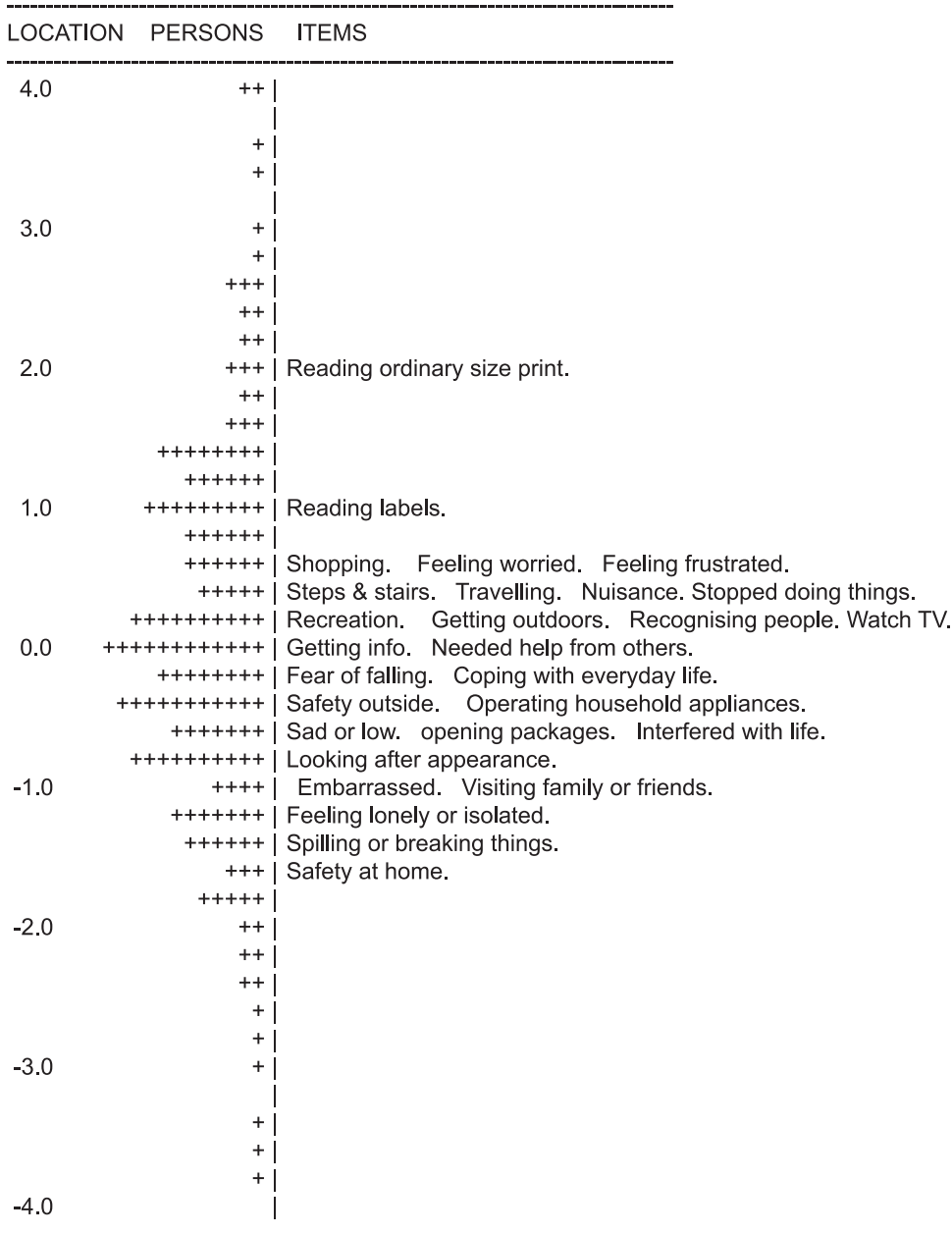


FIGURE 4. Person-item map of the Rasch-scaled IVI questionnaire, showing the distribution of Rasch-calibrated participant scores (left) and item locations (right).

evidence of differential item functioning or failure of items to fit with an overall latent trait of quality of life.

The use of Rasch analysis has enabled a detailed examination of the operation of the IVI scale. The partial credit rating scale model<sup>58</sup> was used to evaluate the ordering of categories (threshold ordering), and the evidence suggests that the response scale of the original version of the IVI was not optimal. The original 32-item IVI used six response categories ranging from *not at all* to *can't do because of eyesight*. Analyses indicated significant overlapping between response categories, suggesting that our participants had difficulty consistently discriminating between response options. This was a problem for the very mild end of the response scale: *not at all* overlapped with *hardly at all*. Similarly, at the severe end of the scale: *a lot* overlapped with *can't do because of eyesight*. After the combination of overlapping categories, further analyses showed that a four-rating category (which could be called *not at all, a little, a fair amount* and *can't do because of eyesight*) were effective for 28 items. A three-rating category was used for the

remaining four items. The reduction to a three- or four-category response scale in the measurement of visual disability is consistent with findings from other studies that have investigated response category utilization.<sup>16,17,19,20,26</sup>

The four items that were removed to achieve fit to the overall model recorded high levels of missing data, poor spread, and considerable skewness and showed deviation from the expected model curves. The inadequate fit to the expected model could be due to variability in the visual ability needed to perform specific activities, such as hobbies, or nonvisual factors like relative interest in sport (i.e., going out to sporting events) or the inherent difficulty of the activity (i.e., reading a sign across the street). It has been suggested that the variability in such items generates a substantial level of noise which contributes little to the measurement characteristics of the scale.<sup>18</sup>

Evidence of substantial construct validity of the Rasch-scaled IVI is supported by the absence of DIF for gender, degree of visual impairment, comorbidity, and effect of comor-

TABLE 3. Principal Component Analysis of the Residuals

Item	First-Factor Loading
Ability to see and enjoy TV	-0.114
Taking part in recreational activities	-0.246
Shopping	<b>-0.477</b>
Reading ordinary size print	<b>-0.366</b>
Visiting friends or family	-0.021
Recognizing or meeting people	<b>-0.390</b>
Getting information that you need	-0.198
Looking after your appearance	-0.239
Opening packaging	<b>-0.346</b>
Reading labels or instructions on medicine	<b>-0.395</b>
Operating household appliances and telephone	<b>-0.465</b>
Getting about outdoors	<b>-0.404</b>
Difficulty avoiding falling or tripping	-0.133
Traveling or using transport	-0.111
Going down steps, stairs, or curbs	-0.259
General safety at home	-0.146
Spilling or breaking things	-0.236
General safety outside of home	-0.126
Stop from doing things	0.162
Need help from other people	0.138
Embarrassed	<b>0.397</b>
Frustrated or annoyed	<b>0.391</b>
Lonely and isolated	<b>0.558</b>
Sad or depressed	<b>0.662</b>
Worried about eyesight getting worse	<b>0.408</b>
Difficulty coping with life	<b>0.501</b>
Feels like a nuisance or a burden	<b>0.423</b>
Vision interferes with life in general	<b>0.471</b>

Data show the first-factor loading with two subsets (positive and negative loading items). Only items with factor loadings greater than  $\pm 0.3$  from each subsets were used (in bold). No significant differences were found between the person estimates of the Rasch-scaled IVI and the eight-item positive (*t*-test;  $P = 0.95$ ) and seven-item negative subsets (*t*-test;  $P = 0.97$ ) supporting a unidimensional construct.

bidity on daily living. Considering the multicultural composition of the Australian population, future studies could provide a cross-cultural validation of the IVI using the DIF function. The test of local independence revealed no evidence of multidimensionality, which provides support for the unidimensionality of the revised IVI. The criterion validity of the IVI was demonstrated by its ability to discriminate significantly between participants with mild, moderate, and severe visual impairment.

The person-item map of the Rasch-scaled IVI shows good targeting of the scale, with no apparent floor or ceiling effect. We found only a few participants who did not have difficulty performing even the most difficult items and others who had substantial difficulty performing the easiest activities, consistent with a sample of visually impaired people attending low-vision rehabilitation. In addition, the good targeting of item difficulty to patients level of participation suggests that the revised IVI is suitable to assess difficulty in performing daily activities across the spectrum of visual disability in individuals living in the community. The person-item map also reveals one of the critical weaknesses of the Likert scoring which assumes that all items are similar in difficulty and all scores of the same worth and can be used in questionnaire development, to ensure accurate targeting.<sup>16,23</sup> For example, reading ordinary size print was identified as a more difficult item to endorse than reading labels or instructions.

The item map also revealed several items representing the same level of difficulty along the ability continuum, suggesting that some items could be removed. However, the revised 28-item IVI is a relatively short questionnaire and, with a

reasonable administration time (mean, 12 minutes), it is unlikely to represent a substantial respondent burden. In addition, it has been argued that low-vision rehabilitation enhances remaining vision for specific activities, and patient-specific information about the effectiveness of the intervention may be lost if items are eliminated.<sup>18,19</sup> For example, although activities like traveling or using transport and going down steps and stairs have the same level of difficulty (0.42 and 0.41 logits, respectively), they are likely to require specific rehabilitation strategies. Although this is not important for the overall IVI score (because this represents a broad underlying construct), the outcome of low-vision rehabilitation with the IVI questionnaire could also be assessed on a question-by-question basis, in which case individual question content would be important. For example, the items relating to traveling and using transport could be used to guide the development of an intervention designed to improve orientation and mobility skills, as well as strategies to deal with a changing environment. Items relating to going down steps and stairs, especially in the home environment, could be used to assess the efficacy of interventions designed at improving the safety in and around the house and obstacle-negotiation strategies. Considering that the sensitivity of the IVI items to change after low-vision rehabilitation has not yet been established, it would be premature to edit the revised questionnaire further.

Our emotional well-being items fitted on the same scale as items measuring difficulty of performing vision specific tasks, which is different from previous findings such as the NEI-VFQ and warrants discussion. The content of a questionnaire determines the latent trait being sampled. If the content is dominated by visual disability items, then the latent trait is visual disability and items particular to other domains are unlikely to fit. This content-determined fit occurs with the NEI-VFQ, which is predominantly a visual disability questionnaire. However, if a questionnaire samples many aspects of quality of life, without the content's being dominated by a particular domain, then the underlying trait is "quality of life." Other quality-of-life questionnaires have

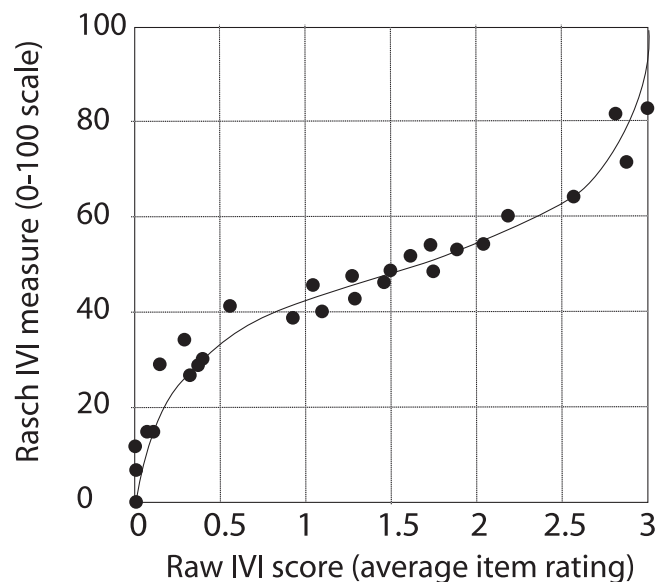


FIGURE 5. Scatter plot of the person measure estimated from Rasch analysis versus the average rating for each person across items (raw IVI questionnaire score). The fit line is generated by double-asymptotic nonlinear regression:  $IVI_{\text{person measure}} = 19.72 \log(IVI_{\text{raw score}} / (3 - IVI_{\text{raw score}})) + 48.29$ .



been shown to include visual disability items that fit with the overall concept—quality-of-life—because they are not a dominant domain (e.g., the Quality of Life Impact of Refractive Correction (QIRC) questionnaire).<sup>45</sup> Items from disability and other domains also fit to the IVI for similar reasons. Although more than half of the items in the IVI involve “task ability” items, these are worded to sample participation and so are not strictly visual disability items. It is this complexity of sampling the impact of visual impairment that makes the IVI a global quality of life measure assessing participation of the visually impaired.

In conclusion, this study demonstrated that the application of the Rasch measurement model supports the revised 28-item IVI as a valid scale for measuring perceived restriction of participation associated with daily living activities, making it suitable for use in assessing the outcomes of low-vision rehabilitation programs. A raw score-to-Rasch person measure conversion is provided to allow other investigators to use the revised IVI without needing to use Rasch analysis. The revised 28-item IVI questionnaire is available on request.

## References

1. U.S. Department of Health and Human Services. *Vision Research—A National Plan: 1999–2003*. Bethesda, MD: National Eye Institute; 1998. NIH Publication 98-4120.
2. Massof RW, Rubin GS. Visual function assessment questionnaires. *Surv Ophthalmol*. 2001;45:531–548.
3. Margolis MK, Coyne K, Kennedy-Martin T, Baker T, Schein O, Revicki DA. Vision-specific instruments for the assessment of health-related quality of life and visual functioning: a literature review. *Pharmacoeconomics*. 2002;20:791–812.
4. de Boer MR, Moll AC, de Vet HC, Terwee CB, Volker-Dieben HJ, van Rens GH. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt*. 2004;24:257–273.
5. Weih LM, Hassell JB, Keeffe JE. Assessment of the impact of vision impairment. *Invest Ophthalmol Vis Sci*. 2002;43:927–935.
6. Lamoureux EL, Hassell JB, Keeffe JE. The determinants of participation in activities of daily living in people with impaired vision. *Am J Ophthalmol*. 2004;137:265–270.
7. Lamoureux EL, Hassell JB, Keeffe JE. The impact of diabetic retinopathy on participation in daily living. *Arch Ophthalmol*. 2004;122:84–88.
8. Hassell JB, Lamoureux EL, Keeffe JE. Impact of age-related macular degeneration on quality of life. *Br J Ophthalmol*. 2006;90:593–596.
9. Noë G, Ferraro J, Lamoureux E, Rait J, Keeffe JE. Associations between glaucomatous visual field loss and participation in activities of daily living. *Clin Exp Ophthalmol*. 2003;31:482–486.
10. Massof RW. The measurement of vision disability. *Optom Vis Sci*. 2002;79:516–552.
11. Fisher WPJ. The Rasch debate: validity and revolution in educational measurement. In: Wilson M, ed. *Objective Measurement: Theory into Practice*. Norwood, NJ: Ablex; 1994:36–72.
12. Tennant A, Penta M, Tesio L, et al. Assessing and adjusting for cross cultural validity of impairment and activity limitation scales through Differential Item Functioning within the framework of the Rasch model: the Pro-ESOR project. *Med Care*. 2004;42:37–48.
13. Wright BD, Linacre JM. Observations are always ordinal, measurements must be interval. *Arch Phys Med Rehabil*. 1989;70:857–860.
14. Fisher WP Jr, Eubanks R, Marier RL. Equating the MOS SF36 and the LSU HSI Physical Functioning Scales. *J Outcome Meas*. 1997;1:329–362.
15. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psych*. February 2, 2006 (online publication in advance of print).
16. 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.
17. 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.
18. Stelmack JA, Szlyk JP, Stelmack TR, et al. Psychometric properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire. *Invest Ophthalmol Vis Sci*. 2004;45:3919–3928.
19. Stelmack JA, Stelmack TR, Massof RW. Measuring low-vision rehabilitation outcomes with the NEI VFQ-25. *Invest Ophthalmol Vis Sci*. 2002;43:2859–2868.
20. Turano KA, Gerasch DR, Stahl JW, Massof RW. Perceived visual ability for independent mobility in persons with retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 1999;40:865–877.
21. 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.
22. 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.
23. Pesudovs K, Garamendi E, Elliott DB. The Quality of Life Impact of Refractive Correction (QIRC) Questionnaire: development and validation. *Optom Vis Sci*. 2004;81:769–777.
24. Smith HJ, Dickinson CM, Cacho I, Reeves BC, Harper RA. A randomized controlled trial to determine the effectiveness of prism spectacles for patients with age-related macular degeneration. *Arch Ophthalmol*. 2005;123:1042–1050.
25. 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.
26. 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.
27. Garamendi E, Pesudovs K, Stevens MJ, Elliott DB. The Refractive Status and Vision Profile: evaluation of psychometric properties and comparison of Rasch and summated Likert-scaling. *Vision Res*. 2006;46:1375–1383.
28. Keeffe JE, Lam D, Cheung A, Dinh T, McCarty CA. Impact of vision impairment on functioning. *Aust NZ J Ophthalmol*. 1998;26(suppl 1):S16–S18.
29. Mangione CM, Phillips RS, Seddon JM, et al. Development of the Activities of Daily Vision Scale: a measure of visual functional status. *Med Care*. 1992;30:1111–1126.
30. Steinberg EP, Tielsch JM, Schein OD, et al. The VF-14: an index of functional impairment in patients with cataract. *Arch Ophthalmol*. 1994;112:630–638.
31. Mangione CM, Berry S, Spritzer K, et al. Identifying the content area for the 51-item National Eye Institute Visual Function Questionnaire: results from focus groups with visually impaired persons. *Arch Ophthalmol*. 1998;116:227–233.
32. Frost NA, Sparrow JM, Durant JS, Donovan JL, Peters TJ, Brookes ST. Development of a questionnaire for measurement of vision-related quality of life. *Ophthalmic Epidemiol*. 1998;5:185–210.
33. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press; 1960.
34. Andrich D. Rating formulation for ordered response categories. *Psychometrika*. 1978;43:561–573.
35. Guttman LA. The basis for Scalogram analysis. In: Stouffer SA, Guttman LA, Suchman FA, Lazarsfeld PF, Star SA, Clausen JA, eds. *Studies in Social Psychology in World War II*. Princeton: Princeton University Press; 1950:60–90. Vol 4. *Measurement and Prediction*.
36. Smith RM. Fit analysis in latent trait measurement models. *J Appl Measure*. 2000;2:199–218.
37. Andrich D, Lyne A, Sheridan B, Luo G. *RUMM 2020*. Perth, WA, Australia: RUMM Laboratory; 2003.
38. Masters G. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47:149–174.

39. Fisher W. Reliability statistics. *Rasch Measurement Transactions*. Chicago, IL: American Educational Resources Association; 1992.
40. Wright B, Master G. *Rating Scale Analysis*. Chicago: MESA Press; 1982.
41. Smith EV. Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas*. 2002;3:205-231.
42. Holland PW, Wainer H. *Differential Item Functioning*. Hilldale, NJ: Lawrence Erlbaum Associates; 1993.
43. Hassell JB, Lamoureux EL, Keeffe JE. Impact of age related macular degeneration on quality of life. *Br J Ophthalmol*. 2006;90:593-596.
44. Massof RW. Application of stochastic measurement models to visual function rating scale questionnaires. *Ophthalmic Epidemiol*. 2005;12:103-124.
45. Pesudovs K, Garamendi E, Elliott DB. The Quality of Life Impact of Refractive Correction (QIRC) Questionnaire: development and validation. *Optom Vis Sci*. 2004;81:769-777.