

An Evaluation of the Reliability and Validity of the Visual Functioning Questionnaire (VF-11) Using Rasch Analysis in an Asian Population

Ecosse L. Lamoureux,^{1,2} Konrad Pesudovs,³ Julian Thumboo,⁴ Seang-Mei Saw,^{5,6} and Tien Y. Wong^{1,2,6}

PURPOSE. To determine whether the VF-11 is a valid scale to measure visual functioning in an Asian population with vision impairment.

METHODS. Participants from the Singapore Malay Eye Study (SiMES) took part. Visual functioning was assessed by using the VF-11 (a modified version of VF-14 for an Asian population). Rasch analysis was performed on 618 participants with presenting visual acuity < 6/12 in the better eye.

RESULTS. Disordered thresholds were initially evident, indicating that the categories were difficult to discriminate and required category collapsing (from 5 to 4) for nine items. The removal of two misfit items related to driving resulted in a fit of the VF-9 data to the Rasch model ($\chi^2 = 50.5$, $df = 27$, $P = 0.005$). There were no more misfit items. The person separation reliability value was 0.82 which demonstrates that the VF-9 has sufficient ability to discriminate between at least two groups of participants with different levels of visual functioning. The VF-9 significantly differentiated patients stratified by visual acuity demonstrating adequate criterion validity. All items were free of differential item functioning, and there was no evidence of multidimensionality. Targeting of person ability and item difficulty was suboptimal, although this is inevitable in a population-based survey where most people would not be disabled.

CONCLUSIONS. Although the Rasch-modified VF-9 scale achieved fit to the Rasch model, its suboptimal targeting suggests that the instrument does not have the range of items to assess the impact of vision impairment across the severity spectrum of vision loss in this population. (*Invest Ophthalmol Vis Sci* 2009;50:2607–2613) DOI:10.1167/iovs.08-2359

From the ¹Centre for Eye Research Australia, the Royal Victorian Eye and Ear Hospital, University of Melbourne, East Melbourne, Australia; ²Singapore Eye Research Institute, Singapore National Eye Centre, Singapore; ³NH&MRC Centre for Clinical Eye Research, Flinders University and Flinders Medical Centre, Adelaide, Australia; the ⁴Department of Rheumatology and Immunology, Singapore General Hospital, Singapore; and the Departments of ⁵Community, Occupational and Family Medicine and ⁶Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

Supported by a Australian National Health and Medical Research Council Public Health Fellowship (EL).

Submitted for publication May 30, 2008; revised November 19 and December 9, 2008; accepted April 3, 2009.

Disclosure: **E.L. Lamoureux**, None; **K. Pesudovs**, None; **J. Thumboo**, None; **S.-M. Saw**, None; **T.Y. Wong**, 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.

Although objective vision impairment measures such as visual acuity are important, a comprehensive assessment of ophthalmic outcomes should also include measurements from the patient's point of view.¹ Patient-centered functioning is usually assessed with questionnaires, and these have shown that impaired vision significantly reduces activities associated with participation in society and religion, mobility, and visually intensive tasks, and increases the need for community help, family support, and nursing home placement.^{2–12}

A major limitation common to several studies reporting on visual functioning or quality of life (QoL) has been the use of a mean or summary score. Summary scoring, termed Likert scoring, allocates an ordinal assignment of a numerical value to a participant's response and assumes a score based on an interval scale. The validity of such summary scores has been questioned by Rasch analysis—usually referred as the Rasch measurement model.^{1,13–17} The model assumes that the probability that a respondent will select an item is a logistic function of the relative distance between the item location and the respondent location on a linear scale. In other words, the probability that a person will affirm a category within an item is a logistic function of the difference between the person's level of, for example, visual functioning and the level of visual functioning necessary to perform this item, and only a function of that difference. Thus, Rasch analysis is taken as a criterion for the structure of the responses that should be satisfied rather than a simple statistical description of the responses. Both person ability and item difficulty must refer to one trait being measured (i.e., visual functioning) which supports the concept of unidimensionality, which is central to the Rasch model. Once the data fit the Rasch model, estimates of measures on an interval scaling are provided^{1,18,19} that improve the accuracy of scoring and remove measurement noise.^{1,18–20} The transformed overall score can then be used in analysis of variance and regression more readily than the raw total score which has floor and ceiling effects.

To date, however, few vision-specific scales have been validated by Rasch analysis, and even fewer questionnaires have been validated with Rasch analysis for use in Asia, the world's largest continent. An estimated 22 million people are blind, and 67 million with low vision currently live in Asia.²¹ The burden of vision loss in Asia is anticipated to increase with an aging population this century and an understating of the impact on visual impairment on daily function using a psychometrically validated scale is needed for clinical and research purposes.

The Singapore Malay Eye Study (SiMES) was undertaken to determine the prevalence and impact of visual impairment and major eye diseases in urban Asian populations. The 11-item Visual Functioning (VF-11) Questionnaire, a modified version of the 14-item Visual Functioning (VF-14) questionnaire,²² was used to determine the impact of compromised vision on visual functioning. In this article, using Rasch analysis, we determined whether the VF-11 possesses the required psychometric characteristics to measure visual functioning in an Asian pop-

TABLE 1. The Original VF-14 Items and the Modified VF-11 Items Used in the Study

VF-14 Items	Modified VF-11 items
1. Difficulty even with glasses reading small print such as labels on medicine bottles, a telephone book, food labels?	1. Difficulty in reading small print in the telephone book even with glasses?
2. Difficulty even with glasses reading a newspaper or a book?	2. Difficulty in reading newspaper size print even with glasses? Not included
3. Difficulty even with glasses reading a large-print book or large-print newspaper or numbers on a telephone?	3. Difficulty in recognizing friends when you meet them while shopping even with glasses?
4. Difficulty even with glasses recognizing people when they are close to you?	4. Difficulty seeing stairs even with glasses?
5. Difficulty even with glasses seeing steps stairs or curbs?	5. Difficulty in reading street signs or shop signs even with glasses? Not included
6. Difficulty even with glasses reading traffic signs street signs or store signs?	6. Difficulty in filling out 4-D or Toto forms even with glasses?
7. Difficulty even with glasses doing find handwork like sewing, knitting, crocheting, carpentry?	7. Difficulty in playing games—chess or cards even with glasses? Not included
8. Difficulty even with glasses writing checks or filling out forms?	8. Difficulty in cooking even with glasses?
9. Difficulty even with glasses playing games such as bingo, dominos, card games, mahjong?	9. Difficulty in watching television even with glasses?
10. Difficulty even with glasses taking part in sports like bowling, handball, tennis, golf?	10. Difficulty in driving during the day because of vision?
11. Difficulty even with glasses cooking?	11. Difficulty in driving at night time because of vision?
12. Difficulty even with glasses watching television?	
13. Difficulty driving during the day because of your vision?	
14. Difficulty driving at night because of your vision?	

ulation living with visual impairment. We specifically assessed the VF-11 for unidimensionality, interval scaling, internal consistency reliability, criterion validity, appropriate targeting of person ability to item difficulty and the absence of differential item functioning for important disease-related factors.

METHODS

Study Population

The SiMES is a population-based cross-sectional study of Malay subjects residing in Singapore; the study procedures have been described elsewhere.^{12,23,24} Briefly, an age-stratified random sampling procedure was used to select Singaporean Malays aged 40 to 80 years.⁸ Of the 4168 eligible participants from the sampling frame, 3280 (78.7%) participated. Sociodemographic and medical data were recorded with a standardized questionnaire that has previously been described.⁸ The study was conducted in accordance with the Declaration of Helsinki. Ethics approval was obtained from the Singapore Eye Research Institute Institutional Review Board. Data for the study include only participants with a presenting visual acuity in the better eye < 6/12 (or 6/12 or better with restricted fields) from any cause.

Visual Functioning

Visual functioning was assessed with the VF-11 which has been used previously and changed to suit the local cultural context—namely for Chinese Singaporeans.²⁵ Item relevancy was decided in a pilot study of the VF-14 in patients who had cataract surgery in Singapore, and the questions were adapted to suit the local cultural context.²⁶ Briefly, 11 visual functioning questions, modified from the VF-14,²² were used to assess the level of difficulty in performing daily activities (Table 1). Nine of the VF-11 scale items were rated on a numeric scale ranging from 0 (no difficulty) to 4 (unable to perform activity). The remaining two driving items had three response options (1, no difficulty; 2, a little difficulty; 3, a great difficulty).

Trained research assistants conducted face-to-face interviews in Malay and/or English. The questionnaire was translated into Malay and back-translated into English by two different interpreters conversant in both Malay and English. During administration, the participant was given a choice of being interviewed in either Malay or English. All interviewers were fluently bilingual. With the participant's consent,

randomly selected interviews were recorded for periodic review by the investigators for quality control purposes.

Rasch Analysis

Rasch analysis was undertaken to determine specific aspects of validity, reliability, and measurement of the VF-11. We used Rasch analysis software (RUMM2020, 2003; RUMM Laboratory, Perth Australia) and the Andrich rating scale to determine whether the data fit the Rasch model.²⁷ Content validity was evaluated by using person and item fit residual statistics, where it is expected that the mean and SD values approximate 0 and 1, respectively. An overall item-trait interaction score (χ^2) with a statistically nonsignificant Bonferroni-adjusted α value indicates fit to the model and that hierarchical ordering of the items (i.e., from difficult to easy) is consistent across all levels of visual functioning. An estimate of overall scale functioning is the person separation reliability (PSR) index.²⁸ This estimate is closely linked to the targeting of the scale, as it differentiates the number of statistically distinct groups of respondents that can be identified on the trait.²⁹ A PSR index of 0.7 determined by the RUMM software is the equivalent of a G value of 1.5, representing the ability to distinguish two distinct strata of person ability.^{29,30} A value of 0.9 is equivalent to a G value of 3, with the ability to distinguish four strata of person ability. PSR indices (similar to Cronbach's α) of ≥ 0.7 are suitable for group use.³¹

We also assessed the disordered thresholds that occur when participants have difficulty discriminating between the response options. A disordered threshold means literally that a category expected to be "harder" than an adjacent category was actually "easier," but often represents interchangeability of categories. Disordered thresholds indicate a lack of the invariance of items, which is the ratio of difficulties between any pair of items, remains constant across the ability levels of respondents. Category collapsing is often the solution to disordered thresholds. Differential item functioning (DIF) testing was undertaken to determine whether different groups within the sample (age, sex, types of eye disease, degree of visual impairment, and comorbidity), despite equal levels of functioning, respond differently to individual items.

Finally, the unidimensionality of the scale is assessed with a principal components analysis of the residuals. This allows for a test of the local independence of the items. This test implies that once the Rasch factor has been taken into account, there should be no further associations between the items other than random ones. The absence of any

TABLE 2. The Personal Characteristics of the 618 Participants with Vision Impairment (Presenting Visual Acuity < 6/12)

Sex	
Male	234 (37.9%)
Female	384 (62.1%)
Age (y)	
Mean \pm SD	65.4 \pm 10.5
40–49	65 (10.5%)
50–59	113 (18.3%)
60–69	153 (24.8%)
70–80	287 (46.4%)
Comorbid condition	
Yes	263 (42.6%)
No	355 (57.4%)
Eye condition	
Cataract	243 (39.3%)
Uncorrected refractive error	302 (48.9%)
Glaucoma	8 (1.3%)
Diabetic retinopathy	19 (3.1%)
Age-related maculopathy	12 (1.9%)
Other eye conditions	34 (5.5%)
Visual acuity (logMAR)	
Mean \pm SD	0.60 \pm 0.17
Mild (<6/12–6/18)	191 (30.9%)
Moderate (<6/18–6/60)	401 (64.9%)
Severe (<6/60)	26 (4.2%)

meaningful pattern in the residuals is deemed to support the assumption of unidimensionality and the evidence of construct validity. Unidimensionality, in addition to the other adequate fit statistics of the data to the Rasch model, would support the specific validity, reliability, and measurement characteristics of the VF-11.³² To maximize the retention of the initial character of the VF-11 a minimalist approach was taken to item and scale changes such that only those changes essential for satisfactory scale functioning were made.

The unit of measurement in Rasch analysis is logits, which is the natural logarithm of the odds of success in choosing a response. Tasks of average difficulty are assigned 0 logits. Tasks with above-average difficulty get a positive logit score, and tasks with below-average difficulty get a negative logit score. Person ability is defined as 0 logits when the respondent has a 50% chance of endorsing an item of an average difficulty if only two response options are available. For polytomous scoring (more than two response options), as used in this study, and for the Andrich model, a logit is defined as the log odds ratio of the probability of responding with category X to the probability of responding with category X – 1.

For a well-targeted scale, a sample size of 108 can give 99% confidence of the person estimate being within ± 0.5 logits.^{33,34} If the scale is not well targeted (i.e., <15% or >85% endorsement rate), as was anticipated in our population-based study with a high number of

individuals with mild visual disability, then the sample size required for accurate estimation increases to 243. Consequently, the sample size of 618 participants in the present study was considered large enough to give good precision, regardless of the targeting of the sample.

RESULTS

Participants' Characteristics

The characteristics of 618 participants with presenting visual acuity < 6/12 in the better eye are shown in Table 2. The mean \pm SD age was 65.4 \pm 10.5 years, and most of the participants were female ($n = 384$; 62.1%). More than two thirds ($n = 263$; 42.6%) reported at least one comorbid condition, and 67.1% ($n = 427$) had moderate or worse presenting visual acuity in the better eye.

Initial examination of the person fit statistics showed that half of them ($n = 311$, 50.3%) were considered to have a pattern of extreme responses. They either reported that they had no difficulty with the nine items or the items were not applicable. Further analysis found that participants with extreme responses had significantly better vision than did those with a nonextreme pattern of responses (logMAR 0.51 vs. 0.64, $P < 0.001$). They also had a significantly greater proportion of participants with undercorrected refractive error (55% vs. 45%, $P < 0.03$) as the cause of decreased visual acuity. Removal of these participants, however, did not improve the overall item-trait interaction statistics. Nonetheless, considering that those with extremes scores did not contribute to the estimates of the item locations, the validation of the VF-11 was essentially undertaken by using data from the remaining 312 participants.

Fit of the VF-11 Data to the Rasch Model

For ease of interpretation of scores of the VF-11 rating scale, the scoring was reversed for the Rasch analysis giving higher scores to participants with high visual ability. The Andrich rating scale was used to determine whether the VF-11 data fitted the Rasch model.²⁷ One rating scale was used for the nine items with a five-response scale and another one with the two driving items with three response options. Rasch analysis of the VF-11 data initially showed a lack of fit to the Rasch model with a significant item-trait interaction probability (total $\chi^2 = 114.4$, $df = 44$; $P = 0.000000$). Examination of the pattern of item thresholds revealed disordered thresholds for the nine items with the five response options but not for the two driving items. As shown in Figure 1, left, the response category 3 was not consistently chosen by the participants (i.e., that the difficulty of a higher threshold was lower than that of its adjacent lower threshold), indicating that response

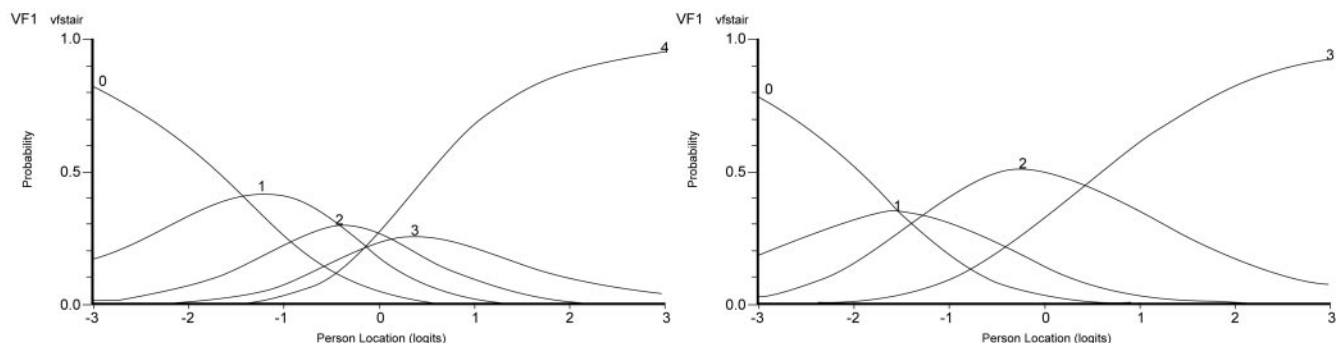


FIGURE 1. *Left:* category probability curves showing disordered thresholds for all items. The response category 3 does not have a range along the ability scale where it is most likely to be chosen. It appears to be interchangeable with categories 2 or 4. *Right:* category probability curves showing ordered thresholds after categories 2 and 3 were collapsed.

TABLE 3. Fit Indices of the VF-9 Scale

Item	Location	Standard Error	Item Fit Residuals	Degrees of Freedom	χ^2	Probability
1. Difficulty seeing stairs	-0.657	0.124	-1.279	260.050	7.749	0.051
2. Difficulty reading street signs	0.535	0.092	0.174	257.490	4.296	0.231
3. Difficulty recognizing friends	0.669	0.089	-0.052	258.340	2.047	0.563
4. Difficulty watching television	-0.295	0.113	-0.162	259.190	4.032	0.258
5. Difficulty cooking	-1.947	0.221	0.015	213.150	1.227	0.746
6. Difficulty playing games	-1.459	0.467	-1.145	34.960	1.356	0.716
7. Difficulty reading newspaper	0.782	0.107	-1.632	186.720	10.590	0.014
8. Difficulty filling out lottery forms	0.889	0.104	-1.127	185.020	10.795	0.009
9. Difficulty reading small print	1.483	0.096	0.303	173.080	3.441	0.328

category 3, *a little difficulty*, was never the most likely response of any participant and was therefore considered a superfluous response option. This result necessitated that categories 2, *moderate difficulty*, and 3, *a little difficulty*, be collapsed (Fig. 1, right), which resulted in ordered thresholds for all items. Consequently, scores for the nine items were recorded from 4, 3, 2, 1, and 0 into 3, 2, 2, 1, and 0.

Rescoring improved the overall item-trait interaction probability (total $\chi^2 = 92.4$, $df = 44$; $P = 0.000027$), although it still remained statistically significant, indicating that the VF-11 did not function within the Rasch model. Examination of the item fit statistics showed that item 11, *difficulty in driving at night time because of vision*, and item 10, *difficulty in driving during the day because of vision*, showed misfit to the model expectation with probabilities of 0.00011 and 0.00026, respectively, which are less than the Bonferroni-adjusted α value of 0.004. These two driving items also recorded the worst response rates (15.5% for both). Misfit items are often indicative that these items do not measure the same construct as other items of the scale. In this particular case, there may be reasons that older patients have difficulty with driving that are not related to their vision (i.e., fear of driving). Item 11 was removed first, followed by item 10.

After item deletion, the VF-9 (since the instrument now has 9 items, not 11, we will hereafter refer to it as the VF-9) showed fit to the Rasch model with a Bonferroni-adjusted nonsignificant item-trait interaction total χ^2 probability ($\chi^2 = 50.5$, $df = 27$, $P = 0.005$). The mean (SD) items and persons fit residuals were -0.57 (0.82) and -0.35 (0.79), respectively, where optimal fit of data to the model would have a mean of 0 and an SD of 1. All items showed fit residuals < 2.5 with a

probability > 0.0045 , indicating no significant deviation from the model (Table 3). The PSR was 0.82, which demonstrates that the VF-9 was able to discriminate between at least two groups of participants with different levels of visual functioning.

Targeting

The person-item threshold map (Fig. 2) shows the persons and items on the same calibrated logit scale for only 307 participants with nonextreme scores. The upper part of the graph is the distribution of persons, and the lower half is the distribution of the thresholds from the items. The map shows a skewed spread of items across the range of respondents' scores suggesting a suboptimal targeting of the patients to the VF-9 item thresholds. Overall, a large number of participants had no difficulty undertaking the hardest items. The mean \pm SD person location logit value (2.07 ± 0.99 logits) substantiates that overall the questionnaire was not optimally targeted and that the participants had a higher level of visual functioning than the average of the scale items (which would be 0 logit).

Overall, the three most difficult items in the VF-9 were *difficulty reading small print* (1.48 logits), *difficulty in filling out lottery forms* (0.89 logits), and *difficulty reading newspaper* (0.78 logits; Table 3). Conversely, the three least difficult items were *difficulty cooking*, *difficulty playing games*, and *difficulty seeing stairs*, with logit scores of -1.95 , -1.46 , and -0.66 , respectively. A similar hierarchical item order has been recorded in prior studies.^{35,36} The response rates for these nine items ranged from 98.5% (watching TV) to 64.4% (playing games).

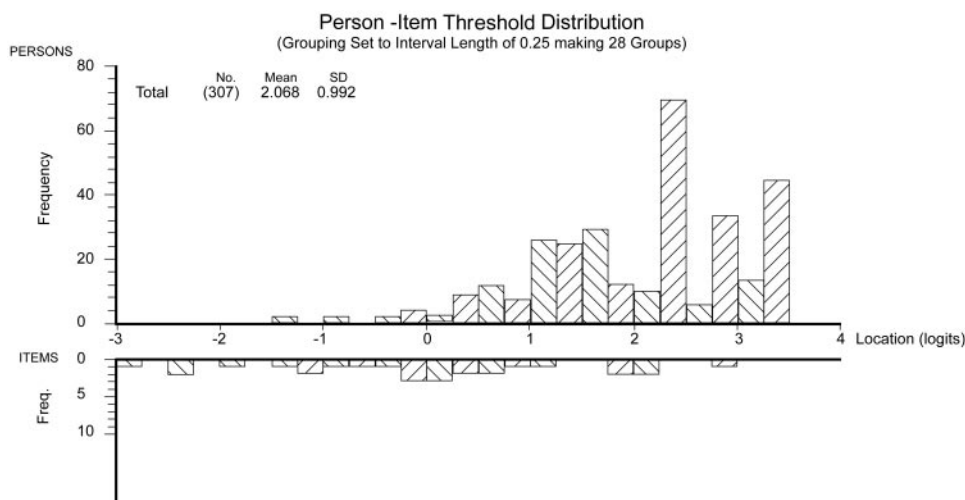


FIGURE 2. The targeting map showing a skewed spread of items across the range of the respondents' scores suggesting a suboptimal targeting of the participants (*top*) to the VF-9 items and thresholds (*bottom*).

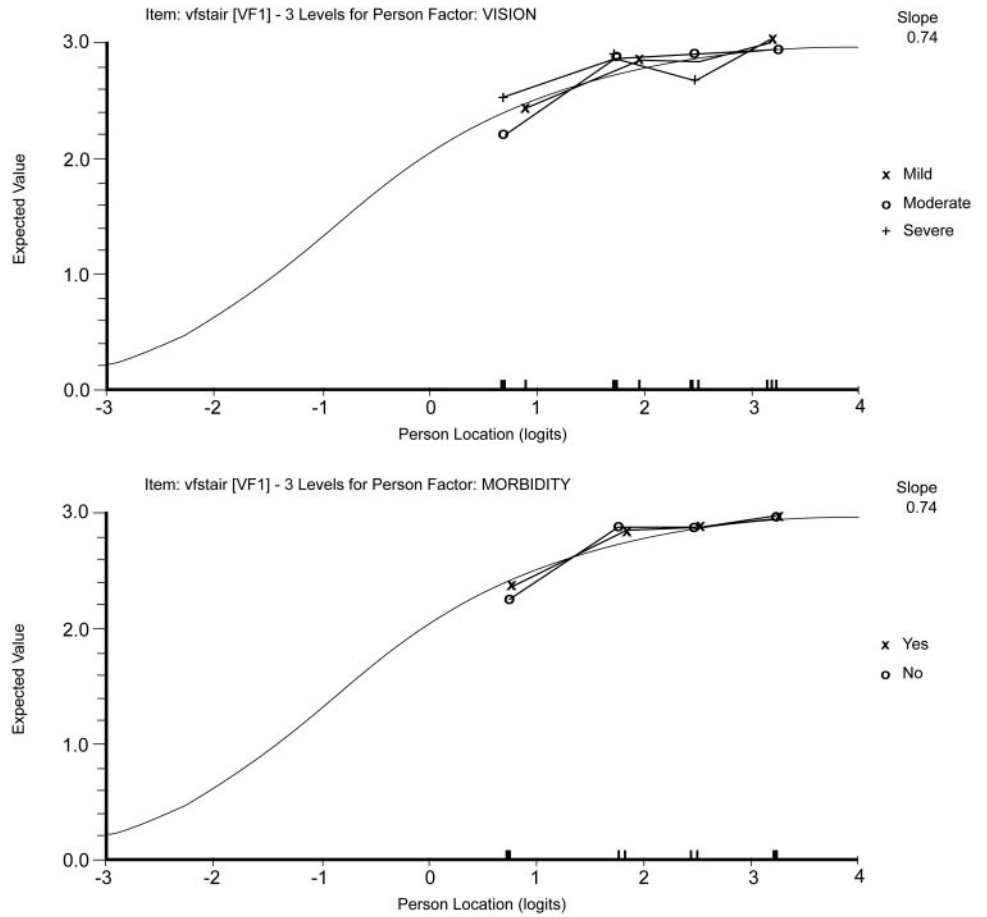


FIGURE 3. Differential item functioning plot for item vfstair, *difficulty seeing stairs*, for (top) vision (mild, moderate, or severe) and (bottom) comorbidity (yes and no).

Differential Item Functioning

DIF was tested for age (40-49, 50-59, 60-69, and 70-80 years); sex (male, female); types of eye disease (cataract, age-related maculopathy; diabetic retinopathy, glaucoma, under-corrected refractive error, and other), degree of visual impairment (mild, moderate, or severe), and comorbidity (yes and no). We found no evidence of DIF on any of these factors for all items of the VF-9. For example, for the item *difficulty seeing stairs*, there was no clear systematic difference in the pattern of responses between those with mild, moderate, or severe vision impairment ($P = 0.82$; Fig. 3, top) or between those with or without other comorbidities ($P = 0.64$; Fig. 3, bottom).

Cultural DIF

To test for culture-specific DIF between Singapore Malays and Western subjects, we compared our VF-9 item measures to previously published VF-14 item measures from patients with low vision in the United States.³⁶ As shown in Figure 4, there was substantial agreement between the two sets of item measures, except for *recognizing friends* ($r = 0.50$ with that item included and $r = 0.79$ with it excluded). The line through the data is the principal component (bivariate regression with orthogonal fit; slope = 1.63 and intercept = 0.21) when *recognizing friends* was excluded. This result suggests strong agreement between U.S. and Singapore Malay patients with low vision in item measure calibration. Population-specific DIF may be evident for the item *recognizing friends*.

Unidimensionality

PCA of the residuals identified two subsets of items consisting of the highest positive (item 8 = 0.75, item 9 = 0.70, and item

7 = 0.59) and negative loading items (item 6 = -0.33, item 3 = -0.57, and item 2 = -0.66). Only 3.36% of estimates were found to be significantly different for these participants. These values are less than the recommended cutoff of 5%, and therefore no evidence of multidimensionality was detected. This finding confirms the internal construct validity of the VF-9 and

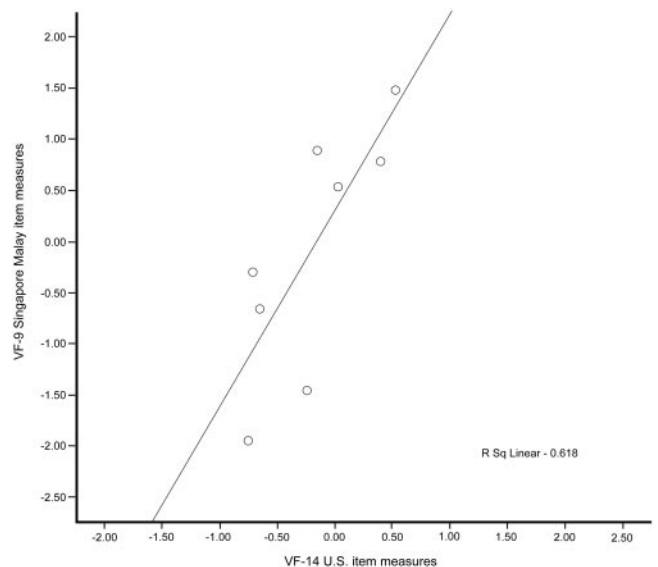


FIGURE 4. Agreement in item measure calibration between U.S. and Singapore Malay patients with low vision for the VF-9 scale except for the item *recognizing friends* ($R^2 = 0.62$).

that the instrument measuring the underlying trait (vision function) that it purports to measure. These results collectively show that the VF-9 is a unidimensional scale and the overall score conforms to an interval scaling.

Criterion Validity

The criterion validity of the Rasch-scaled VF-9 was tested by assessing its ability to discriminate between participants with mild ($VA < 6/12-6/18$), moderate ($<6/18-6/60$), or severe ($<6/60$) vision loss. There was a significant difference between the three groups (ANOVA; $F_{(2, 304)} = 5.73$; $P = 0.003$) with poorer visual acuity being associated with worse visual functioning (mean of 2.32, 2.00, and 1.55 logits for mild, moderate, and severe visual impairment, respectively).

Scoring of the VF-9 Questionnaire

Other investigators wishing to use the VF-9 can use the validation data detailed in this article to convert raw scores into Rasch person measures without having to perform Rasch analysis. We have generated a Rasch scoring key in a spreadsheet (Excel; Microsoft, Redmond, WA; available on request from ecosse@unimelb.edu.au) that converts category responses into 36 item-category calibrations, to create Rasch measurement estimates.

DISCUSSION

In this article, using Rasch analysis, we investigated whether the original VF-11 possesses the required psychometric characteristics to measure visual functioning in an Asian population of Singaporean Malays with low vision. There was evidence of disordered thresholds initially, which necessitated that the response options be collapsed from five to four categories for 9 of the 11 items. In addition, the two "driving" questions were removed from the scale, as they showed evidence of misfit. Subsequently, the remaining nine items were found to fit the Rasch model which indicates that the VF-9 is an appropriate scale to measure visual functioning in this population. We demonstrated that the VF-9 is a unidimensional scale and is free of DIF for several important disease related factors, including culture. The scale possesses interval scaling and good person separation reliability. There was evidence, however, of suboptimal targeting of person ability to item difficulty which suggests that further work is needed to optimize the performance of the instrument.

There have been previous attempts to reduce the original VF-14 questionnaire to fewer items. Uusitalo et al.³⁷ created a shortened instrument, the VF-7, by analyzing the results of the VF-14 administered to patients with cataract and then selecting seven items that correlated best with patient satisfaction. Similarly, to improve efficiency, Mallinson et al.³⁵ demonstrated the use of Rasch measurement in six subsets of seven items of the original VF-14. Although our initial choice of 11 items was motivated by items that were considered relevant to the Singaporean culture, examination of our PSR index suggests that the VF-9 did not lose measurement precision in the shortened version. Our PSR index was 0.82, which indicates that the scale can distinguish between several distinct strata of person ability and has good reliability. In addition, this value compares favorably with that of the VF-14 (0.84) and indicates that the two scales are comparable and the shortening of the original VF-14 did not affect the measurement precision of the VF-11.^{35,36} Similarly, we found that the items *difficulty reading small print* and *difficulty cooking* were considered to be the most and least difficult items, respectively, on the VF-9. Similar findings were noted after the VF-14 underwent Rasch analysis,

which indicates the comparability of the item measures of the two scales.^{35,36}

Although it is difficult to compare vision-specific areas of activity limitation between Singaporean Malays and other populations because of the use of different scales and psychometric evaluation techniques, it can be hypothesized that activities associated with near and distance reading are consistently rated as the most difficult to undertake. For example, in Australians with low vision, it was also found that after Rasch analysis of the Impact of Vision Impairment (IVI) Questionnaire, the two most difficult items were *reading ordinary size print* and *reading labels or instructions on medicines*.¹⁴ Similarly, after Rasch analysis of the Activities of Daily Vision Scale (ADVS) in British patients awaiting cataract surgery *reading signs* was found to be the most challenging activity.³⁸ A parallel result was found in people with visual impairment living in East Timor who rated the most difficult item of the Rasch-analyzed Timor-Leste vision-specific QoL scale as being the *reading things such as a newspaper, bible, or book*.³⁹ These findings suggest that difficulty associated with reading activities is consistent across people with low vision, irrespective of their cultural differences. Our findings also highlight the need for the development and validation of an item bank that would enhance our ability to compare visual functioning and quality of life across cultures and significantly open up the field of low vision to international comparison studies.

The application of Rasch analysis to the VF-9 has allowed greater scrutiny of the performance of the response scale that would not have been possible with the traditional approach to test development (i.e., classic test theory). Singaporean Malays with low vision consistently did not use the response option *a little difficulty* on any of the nine items of the VF-9. The nonuse of this option necessitated the collapsing of the response category with the adjacent one (i.e., *moderate difficulty*), which subsequently produced a consistent endorsement of all response options across items. The nonutilization of categories is not unusual in scales with many response options, or when the labeling of options is too similar to each other, which can be confusing or open to misinterpretation. Our finding is consistent with other vision-specific scales that have benefited from a shortening of their response scales after Rasch analysis.^{14,38,40-42}

The person-item threshold map showed that numerous participants recorded a positive person logit score, which suggests that the person's level of functioning is higher than the mean required level of functioning for the items. Considering that only a small percentage of participants (4.2%) were considered to have severely impaired vision, it is possible that our low-vision sample was not impaired enough to experience substantial visual functioning difficulties. On the other hand, it could be argued that since almost two thirds (64.9%) of our sample were considered to have moderate vision impairment (categorized as presenting visual acuity ranging from $<6/18$ to $6/60$), a better targeting was anticipated. Perhaps the reason for this result is the inclusion of refractive error as the dominant cause of poor visual acuity, because low myopia may cause a decrease in VA, but enable good reading vision, and thus may not cause as much visual disability as cataract or macular degeneration. Regardless, the poor targeting of VF-9 may indicate the instrument does not include the content required to fully assess the impact of vision impairment across the severity spectrum of vision loss in this population. Future investigations are therefore needed to determine whether the inclusion of items of greater difficulty improves the targeting of patients with low vision in this population.

In conclusion, there are very little published data on vision-specific QoL or vision functioning scales in Asia that have been validated with Rasch analysis, which is increasingly considered

the gold standard of modern psychometric scale validation. The Rasch-scaled VF-9 in this Singaporean Malay population demonstrated internal consistency and criterion validity, with no evidence of DIF and multidimensionality. However, targeting was suboptimal, suggesting that the VF-9 is not suitable for population-based studies and that additional items are needed to optimize the instrument's performance.

References

- Pesudovs K. Patient-centered measurement in ophthalmology: a paradigm shift. *BMC Ophthalmol.* 2006;6(1):25.
- Klein BE, Klein R, Lee KE, Cruickshanks KJ. Performance-based and self-assessed measures of visual function as related to history of falls, hip fractures, and measured gait time. The Beaver Dam Eye Study. *Ophthalmology.* 1998;105(1):160-164.
- Rovner BW, Ganguli M. Depression and disability associated with impaired vision: the MoVies Project. *J Am Geriatr Soc.* 1998;46(5):617-619.
- McCarty CA, Nanjan MB, Taylor HR. Vision impairment predicts 5 year mortality. *Br J Ophthalmol.* 2001;85(3):322-326.
- Wang JJ, Mitchell P, Cumming RG, Smith W. Visual impairment and nursing home placement in older Australians: the Blue Mountains Eye Study. *Ophthalmic Epidemiol.* 2003;10(1):3-13.
- West SK, Munoz B, Rubin GS, et al. Function and visual impairment in a population-based study of older adults. The SEE project. *Invest Ophthalmol Vis Sci.* 1997;38(1):72-82.
- Wang J, Mitchell P, Smith W, Leeder SR. Factors associated with use of community support services in an older Australian population. *Aust N Z J Public Health.* 1999;23(2):147-153.
- 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(2):265-270.
- Lamoureux EL, Hassell JB, Keeffe JE. The impact of diabetic retinopathy on participation in daily living. *Arch Ophthalmol.* 2004;122:84-88.
- Weih LM, Hassell JB, Keeffe JE. Assessment of the impact of vision impairment. *Invest Ophthalmol Vis Sci.* 2002;43:927-935.
- Hassell JB, Lamoureux EL, Keeffe JE. Impact of age related macular degeneration on quality of life. *Br J Ophthalmol.* 2006;90(5):593-596.
- Lamoureux EL, Chong E, Wang JJ, et al. Visual impairment, causes of vision loss, and falls: the Singapore Malay Eye Study. *Invest Ophthalmol Vis Sci.* 2008;49(2):528-533.
- Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests.* Chicago: University of Chicago; 1960.
- Lamoureux E, Pallant J, Pesudovs K, et al. The Impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci.* 2006;47:4732-4741.
- Wright BD, Linacre JM. Observations are always ordinal, measurements must be interval. *Arch Phys Med Rehabil.* 1989;70:857-860.
- 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.
- Massof RW. The measurement of vision disability. *Optom Vis Sci.* 2002;79:516-552.
- 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.
- Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C. Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care.* 2004;42(1 Suppl):I25-I36.
- 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.
- Resnikoff SPD, Pascolini D, Etya'ale D, et al. Global data on visual impairment in the year 2002. *Bull World Health Organ.* 2004;82(11):844-851.
- Steinberg EP, Tielsch JM, Schein OD, et al. The VF-14. An index of functional impairment in patients with cataract. *Arch Ophthalmol.* 1994;112(5):630-638.
- Foong AW, Saw SM, Loo JL, et al. Rationale and methodology for a population-based study of eye diseases in Malay people: The Singapore Malay eye study (SiMES). *Ophthalmic Epidemiol.* 2007;14(1):25-35.
- Wong T, Chong E, Wong W, et al. Prevalence and causes of visual impairment and blindness in an urban Malay Population: the Singapore Malay Eye Study (SiMES). *Arch Ophthalmol.* 2008;126(8):1091-1099.
- Saw SM, Foster PJ, Gazzard G, Seah S. Causes of blindness, low vision, and questionnaire-assessed poor visual function in Singaporean Chinese adults: The Tanjong Pagar Survey. *Ophthalmology.* 2004;111(6):1161-1168.
- Lamoureux EL, Chong EW, Thumboo J, et al. Vision impairment, ocular conditions, and vision-specific function: The Singapore Malay Eye Study. *Ophthalmology.* 2008;115:1973-1981.
- Andrich D. Rating formulation for ordered response categories. *Psychometrika.* 1978;43:561-573.
- Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Med Care.* 2004;42(1Suppl):I17-I24.
- Fisher WP Jr. Reliability statistics. *Rasch Measure Trans.* 1992;6:238.
- Wright B, Master G. *Rating Scale Analysis.* Chicago: MESA Press; 1982.
- Nunnally CJ. *Psychometric Theory.* New York: McGraw-Hill; 1978.
- Lundgren-Nilsson A, Grimby G, Ring H, et al. Cross-cultural validity of functional independence measure items in stroke: a study using Rasch analysis. *J Rehabil Med.* 2005;37(1):23-31.
- Linacre J. Sample size and item calibration stability. *Rasch Measure Trans.* 1994;7:28.
- Pallant JF, Miller RL, Tennant A. Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis. *BMC Psychiatry.* 2006;6:28.
- Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Med Care.* 2004;42(suppl 1):I17-I24.
- Massof RW. An interval-scaled scoring algorithm for visual function questionnaires. *Optom Vis Sci.* 2007;84(8):689-704.
- Uusitalo RJ, Brans T, Pessi T, Tarkkanen A. Evaluating cataract surgery gains by assessing patients' quality of life using the VF-7. *J Cataract Refract Surg.* 1999;25(7):989-994.
- 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.
- du Toit R, Palagyi A, Ramke J, et al. Development and validation of a vision-specific quality-of-life questionnaire for Timor-Leste. *Invest Ophthalmol Vis Sci.* 2008;49(10):4284-4289.
- 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.
- 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(11):3919-3928.
- Lamoureux EL, Pallant JF, Pesudovs K, et al. The impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci.* 2006;47(11):4732-4741.