1040-5488/09/8607-0836/0 VOL. 86, NO. 7, PP. E836–E844
OPTOMETRY AND VISION SCIENCE

ORIGINAL ARTICLE

Rasch Analysis of the Quality of Life and Vision Function Questionnaire

Vijaya K. Gothwal*, Thomas A. Wright†, Ecosse L. Lamoureux*, and Konrad Pesudovs‡

ABSTRACT

Purpose. The Quality of life and Vision Function Questionnaire (QOL-VFQ) was developed using classical test theory to assess outcomes of cataract surgery. The aim of this study was to examine the psychometric properties of this questionnaire using Rasch analysis in a cataract population.

Methods. The 17-item QOL-VFQ was self-administered to 389 patients waiting to undergo cataract surgery. The QOL-VFQ and its five subscales were assessed for fit to the Rasch model. Rasch analysis was used to estimate interval level measures of "visual ability" from ordinal scores for the QOL-VFQ and its five subscales. Unidimensionality, item fit, response category performance, and targeting of items to patients were assessed.

Results. The QOL-VFQ and its five subscales showed ordered category thresholds. Despite removal of two misfitting items the person separation reliability was high and the QOL-VFQ could distinguish among three strata of patient ability. However, there was suboptimal targeting of patient ability to item difficulty as most of the patients had higher levels of visual functioning. None of the subscales demonstrated acceptable psychometric properties.

Conclusions. These results support the good overall functioning of the QOL-VFQ in patients with cataract. However, adding more items, that suit the more able patients including those who are awaiting cataract surgery in the fellow eye, will help improve the targeting.

(Optom Vis Sci 2009;86:E836–E844)

Key Words: Quality of life and Vision Function questionnaire, cataract, Rasch analysis

There is growing consensus among researchers that patient-reported outcomes (or questionnaires) are important in the comprehensive assessment of patients with ocular conditions including cataract, especially for the measurement of outcomes.[1–6] Questionnaires can provide a comprehensive picture of visual disability beyond the results of a clinical evaluation.[7] Using classical test theory (CTT),[8–10] the Quality of Life and Vision Function Questionnaire[11] (QOL-VFQ) was developed and validated in a heterogeneous group of ocular patients, which included a subset with cataract. A major limitation of the CTT approach is the use of Likert scoring with the erroneous allocation of equal weight to all the items in the questionnaire, and thereby treating the whole questionnaire as interval scale based on ordinal level scoring.[10,12–15] This limitation is overcome by the use of item response theory,[16,17] in particular Rasch analysis, that makes use of interval-level data.[15,18] Recently, Rasch analysis has been applied to revalidate the other visual disability questionnaires, Visual Functioning-14 (VF-14), Activities of Daily Vision Scale (ADVS), and Visual Disability Assessment for the cataract population.[3,6,19]

The QOL-VFQ was reported as reliable and valid questionnaire, by CTT, for the assessment of "self-reported visual satisfaction by ophthalmic patients with visual impairment from chronic eye diseases" including cataract, glaucoma, age-related macular degeneration, branch retinal vein occlusion, and minor refractive errors (<5 D).[11] The limitations of CTT and their potential resolution with Rasch Analysis prompted us to conduct this study. The aim of this study was twofold: first, to determine whether item response theory scoring criteria, using Rasch analysis would be appropriate for the 17-item QOL-VFQ in patients with cataract; and second to provide the clinicians and researchers with an Excel sheet for ready conversion of raw scores to Rasch scores, and so as to obviate the need for Rasch Analysis for routine clinical outcomes assessment.

*PhD
†BPsyc(Hons)
‡PhD, FAAO
Department of Ophthalmology, NH&MRC Centre for Clinical Eye Research, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, South Australia, Australia (VKG, TW, KP), Meera and L. B. Deshpande Centre for Sight Enhancement, L. V. Prasad Eye Institute, Hyderabad, India (VKG), Department of Ophthalmology, Centre for Eye Research Australia, The University of Melbourne, Melbourne, Victoria, Australia (ELL), Vision CRC, Sydney, New South Wales, Australia (ELL), and Singapore Eye Research Institute, Singapore National Eye Centre, Singapore (ELL).

METHODS

Quality of Life and Vision Function Questionnaire

The Quality of Life Visual Functioning questionnaire was described as being able to measure "quality of life, visual functioning, and visual satisfaction."[11] The QOL-VFQ consists of 17 questions (Table 1) that can be grouped into six subgroups; self-assessment of visual satisfaction, visual field, distance visual acuity, near visual acuity, sensory adaptation (light/dark adaptation, glare), and color vision. Each question is rated on a three-point scale consisting of (1) "not at all," (2) "quite a lot," and (3) "very much." Therefore, higher scores represent poorer visual functioning. The total questionnaire is calculated by averaging the six subgroup scores.

Three preliminary questions that intend to explore the patient's attitude with regard to his/her health condition did not contribute to the overall questionnaire. Therefore, these questions were not included in our analyses. Notably, the QOL-VFQ was developed for an Italian population. The instrument was reported in English, and we used this version of the QOL-VFQ for this study. Also, the instrument was originally interviewer administered but in this study it was self-administered.

Subjects

Patients awaiting a cataract extraction procedure in one eye or both eyes at the Flinders Eye Centre, Flinders Medical Centre, Adelaide, South Australia comprised the subjects of this study. Patients were

TABLE 1.
Items included in the 17-item Quality of Life and Vision Function Questionnaire

Item no.	Description	Subscale
[a]A. In general would you say your health is		
[a]B. Do you think you get sick more frequently than other persons?		
[a]C. Are you pessimistic regarding your health?		
1.	Does quality of vision prevent you from performing daily activities normally?	Overall self-assessment of visual satisfaction
2.	Are you unhappy about your visual condition?	Overall self-assessment of visual satisfaction
3.	How much are you concerned with a possible worsening	Overall self-assessment of visual satisfaction
4.	Because of your visual problems do you feel less inclined to meet people/friends/relatives?	Overall self-assessment of visual satisfaction
5.	Because of your visual problems do you feel useless or a burden to others?	Overall self-assessment of visual satisfaction
6.	Because of your vision do you have problems crossing a street?	Self-assessment of visual field
7.	Because of your vision do you bump against other people when in crowded areas?	Self-assessment of visual field
8.	Because of your vision do you have problems in perceiving a dip on the ground or step?	Self-assessment of visual field
9.	Because of your vision do you have problems in reading prices in a shop window?	Self-assessment of distance visual acuity
10.	Because of your vision do you have problems in recognizing people across the street?	Self-assessment of distance visual acuity
11.	Because of your vision do you have problems in recognizing a person in a crowded room?	Self-assessment of distance visual acuity
12.	Because of your vision do you have problems in reading an article in a newspaper or names/numbers in the telephone directory?	Self-assessment of near visual acuity
13.	Because of your vision do you have problems in doing a manual activity such as cooking, sewing, cutting your nails?	Self-assessment of near visual acuity
14.	Does your vision deteriorate in bright light?	Self-assessment of sensory adaptation (light-dark adaptation, glare)
15.	Does your vision deteriorate in dim light?	Self-assessment of sensory adaptation (light-dark adaptation, glare)
16.	Do you have a driving license? If "YES": how much is your driving disturbed by the lights of oncoming cars? If "NO": how much is your vision disturbed by the lights of oncoming cars?	Self-assessment of sensory adaptation (light-dark adaptation, glare)
17.	How much problem do you have in recognizing colors?	Self assessment of color vision

[a]Preliminary questions.

aged 18 years or older, English speaking and had no severe cognitive impairment. Although on the cataract surgery waiting list (patients wait for an average of 3 to 4 months for surgery) patients were sent the QOL-VFQ for self-completion and return via a self-addressed envelope. Ethical approval was obtained, and all patients who agreed to participate signed a consent form. The study was conducted in accordance of the Declaration of Helsinki.

These patients had coexisting ocular and systemic comorbidities and this study sample appears to be representative of the elderly cataract population in Australia.[20] Mean age of the patients was 74.2 years (SD, 9.6). The male/female ratio was 4:5. The socio-demographic data of the patients who completed the QOL-VFQ is shown in Table 2.

Clinical Assessment

Routine clinical assessments were performed before cataract extraction. Visual acuity assessments were performed using computerized testing based on logMAR principles and the illumination was 150 cd/m^2.

All the assessments were performed binocularly as binocular acuity is thought to be representative of real world ability.[21,22] Thus, we used binocular visual acuity in all our analysis.

TABLE 2.
Sociodemographics of the study population for QOL-VFQ (N = 389)

Characteristic	Result
Mean age (y) ± SD	74.2 ± 9.6
Gender, n (%)	
Male	173 (44.5)
Female	216 (55.5)
Binocular visual acuity	
Mean ± SD	
LogMAR	0.22 ± .20
Snellen	6/9.5^{-1}
Range	
LogMAR	−0.26 to 1.00
Snellen	6/3^{-2} to 6/60
Awaiting second-eye surgery, n (%)	160 (42.7)
Ocular comorbidity[a], n (%)	
Present	187 (49.2)
Absent	193 (50.8)
Duration of cataract (y)	
Median	1
Range	0–80
Systemic comorbidity[b], n (%)	
Present	314 (89.5)
Absent	37 (10.5)

[a]Includes glaucoma, diabetic retinopathy, age-related macular degeneration etc., and six cases had missing data.
[b]Includes diabetes, hypertension, angina etc., and 38 cases had missing data.

Rasch Analysis

Over the last decade, a number of studies that have used Rasch analysis and have dealt with it in detail (refer to ref. 7 for a detailed mathematical description of the Rasch model). A major advantage of Rasch analysis is that it places the items and the persons along a single ratio scale, the units being logits (i.e., log odds). For ease of interpretation and consistent with the idea that higher scores represent better functioning,[23,24] we reversed the rating scale for Rasch analysis. In this study, logits of higher positive magnitude represent greater item difficulty and higher person ability.

In brief, the Rasch model is a probabilistic model meaning that a certain amount of error is assumed in the model and the use of the model will, therefore, be driven by its goodness-of-fit of model to the data. Thus, there are three main parameters that are assessed in this analysis: behavior of the rating scales (or category thresholds), fit statistics, and separation indices.

The Rasch rating scale model is developed for use with data with ordered response categories. Therefore, the QOL-VFQ data were fitted to the Rasch rating scale model using the Winsteps software[25] (ver. 3.66, Chicago, Winsteps.com). Under this model, category thresholds are the same for all the items in a scale, and the items differ only in their locations whereby the thresholds are relative to item difficulty. The category probability curves provided by Winsteps are a graphical representation of the usage of the response categories that enables the researcher to decide if usage is satisfactory or whether changes need to be made. Response scale malfunction may require repair such as a change in the number or description of response categories. Where response categories are disordered or underutilized, a simple solution can be to combine or collapse adjacent response categories. Collapsing categories can often improve the performance of the overall scale, as well as reduce burden on the respondent and save time. However, the new response categories remain theoretical. The new response categories should be tested for appropriate functioning before the redesigned questionnaire can be recommended for future use.[26,27]

Key indicators of questionnaire functioning within the Rasch model are person separation statistics. Separation indices include the person separation index and reliability. The higher the reliability, the better the questionnaire is in terms of its ability to discriminate among subjects. Values of a person separation of 2.0 results in a reliability of >0.8 and is the minimum recommended acceptance level.[28]

The two fit statistics that are used include the Infit and Outfit. Infit represents the information-weighted mean square residual difference between the observed and expected responses. The Outfit statistic is the unweighted mean square residual and is more sensitive to outliers. When both statistics have a value of 1 satisfactory model fit is suggested. Values <0.7 (indicate redundancy) and >1.3 (indicate inconsistent responses) suggest misfit of the model and the data. A misfitting item indicates that the item is either poorly defined or is measuring something different (for e.g., different ability).[6,29,30] Such misfitting items can either be revised to improve clarity or the response option clarified to the subject.[31] In addition, there may be some odd responses causing misfit to the model. However, if the items still misfit then they should be deleted and while deleting it is more important to delete the underfitting ones (>1.3) in preference to the overfitting ones. Retaining the overfitting ones only increases the length of the

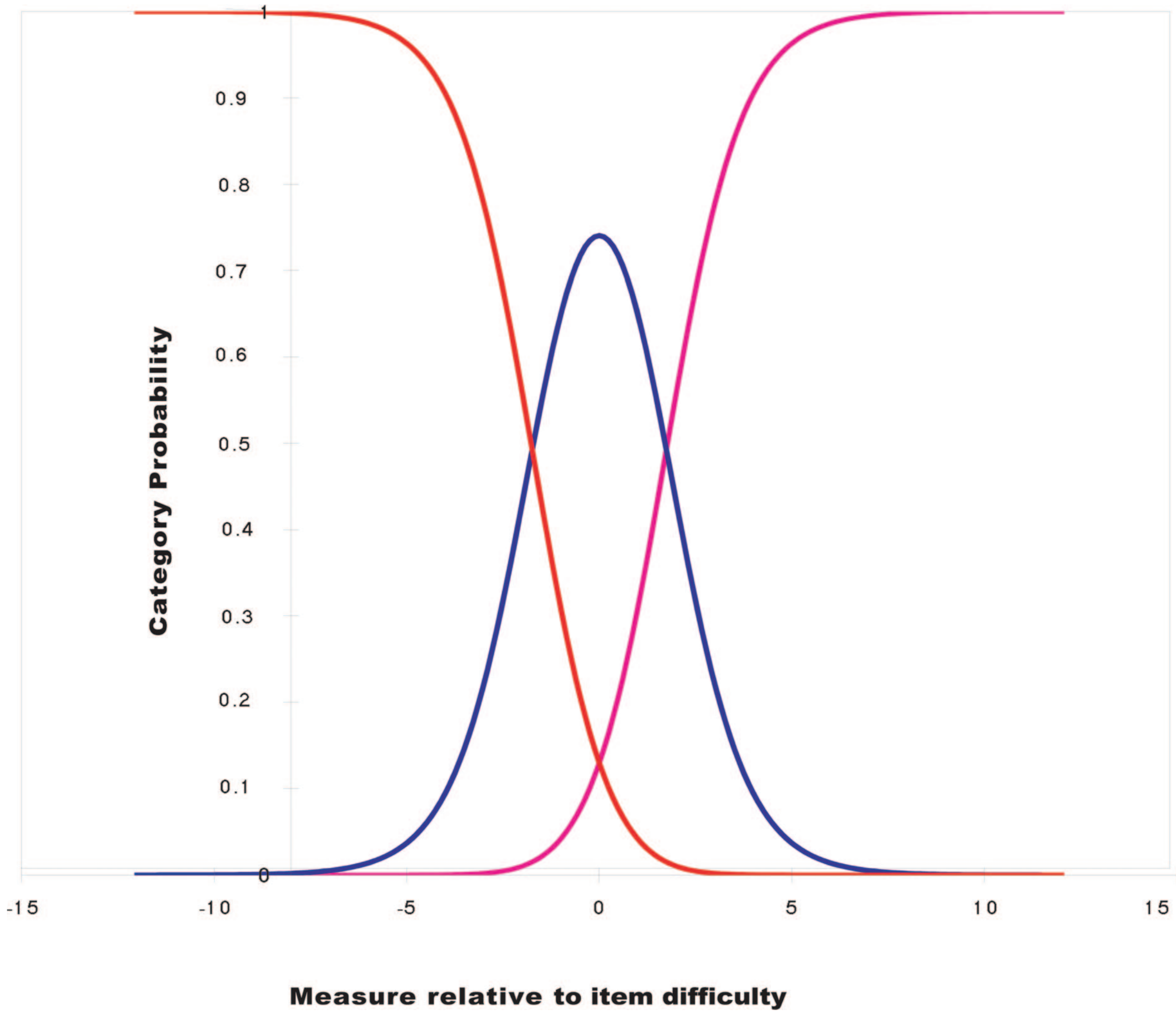

FIGURE 1.

Category probability curves for the 15-item QOL-VFQ illustrating the range over which each of the three categories is most likely to be chosen. Boundaries occur at points along the scale where the category most likely to be chosen changes from one to the next.

questionnaire but does not do much harm to the measurement properties of the questionnaire.

The other important attribute of items that is tested with Rasch analysis is targeting of item difficulty to person ability.[32] A good measure should be able to differentiate persons across the full spectrum of ability in the population. Therefore, the instrument requires items to vary in difficulty over the range of person ability. Rasch analysis illustrates targeting of item difficulty to patient ability in the person-item map and reports the difference in targeting between items and patients in terms of the difference in means. Ideal targeting exists when the mean of items equals the mean of persons; the greater the difference in means the poorer is the targeting.

One of the major requirements for Rasch analysis is unidimensionality and local independence.[33] Unidimensionality means that the questionnaire is measuring a single latent trait or ability, and local independence means that a subject's responses to the items are statistically independent. We assessed for these properties by performing a principal components analysis of the residuals. We used the criterion that the first contrast should have an eigenvalue of >2.0 to cause us to reject the assumption of unidimensionality, as this level was greater than the magnitude seen with random data.[25]

We also assessed Differential Item Functioning (DIF) which occurs when subjects with equal amount of the latent trait (visual disability in this case) respond differently to a particular item. We assessed DIF by age, gender, presence or absence of systemic and ocular comorbidities, cataract status (i.e., if the patient underwent cataract surgery in any eye). Cataract surgery is performed over a wide range of ages, DIF by age in the groupings <74 and ≥74 years was conducted to explore whether it had a confounding effect on item responses. Similarly, DIF by gender was assessed as some items may be gender-specific. DIF by cataract status was assessed as it is possible that tasks difficulty varies depending on whether vision is poor in one eye or two. Similarly, some items may be relatively more difficult in the presence of coexisting systemic or ocular co-morbidities, and therefore it was decided to assess DIF for these variables too. We used the following criteria for DIF assessment: small or absent if the difference in logits was <0.50 logits, 0.50 to 1.0 logits as minimal (but probably inconsequential) DIF and >1.0 logit as notable DIF. If significant and meaningful DIF is found, it may indicate that the interpretation of the scale may differ by group and is, therefore, biased by gender or age etc.

We analyzed the subscales separately and used the same procedures and criteria for reliability and validity as that of the overall questionnaire. However, the subscale for color vision contains only one item and Rasch analysis cannot be done for a single item.

We used SPSS software (version 15.0, Chicago, IL) to perform all the descriptive analysis. We set statistical significance at $p < 0.05$.

RESULTS

Psychometric Properties of the QOL-VFQ

Rating Scale and Reliability

A total of 389 patients were included in this study. The missing item responses ranged between 0.08 and 5.7%. The category thresholds were all ordered (Fig. 1) indicating the original rating scale functioned well. The overall performance of the QOL-VFQ was acceptable with satisfactory fit to the Rasch model. High measurement reliability was demonstrated by high separation indices for 17-item QOL-VFQ indicating the capacity to distinguish between three strata of patient ability (Table 3).

Item fit

Two items misfit—"How much is your driving/vision disturbed by the lights of oncoming cars"? and "How much problem do you have in recognizing colors"? Of these, the former was most misfitting (Infit MNSQ, 1.33) and after it was deleted, the item related to recognizing colors misfit (Infit MNSQ, 1.31). After deletion of these items, the remaining 15 items fit well (Table 4), and the separation indices continued to remain high (Table 3). The category thresholds were all ordered for the 15-item QOL-VFQ.

Targeting

Fig. 2 represents the person-item map for the revised 15-item QOL-VFQ that consists of the person's ability level (left) and item difficulty (right). The person ability ranged from −6.52 to 6.52 logits with a mean of 2.44 logits and a standard deviation of 2.32 logits. The range of person ability was not significantly different from a normal distribution (K-S test = 0.96, p = 0.31). There was a large ceiling effect with one-half persons located above the mean value of the most difficult item—read newspaper. The person-item map demonstrated an uneven spread of items across the entire range of person ability, suggesting that there was less than optimal targeting of item difficulty to patient ability. Most of the items lined up with the bottom one-half persons, suggesting that most of these items were targeting the less able patients, and, therefore, patients with higher visual ability (or less impaired) could not be differentiated well by the items; especially those at the ceiling. The positive mean person ability suggested that the patients in our study had a higher level of visual functioning than the average difficulty of the QOL-VFQ items.

Some of the more difficult items (in order of difficulty) were related to reading an article in newspaper, concern with possible worsening of one's visual condition and being unhappy with one's visual condition. In contrast, some of the less difficult items (in order of ease) were related to feeling less inclined to meet people/friends, feeling useless or burden to others, and bump against other people when in crowded areas (Table 4).

Differential Item Functioning

The QOL-VFQ showed DIF by gender for five items. Four of these showed moderate DIF and only one showed large DIF. The four items that showed moderate DIF and were rated relatively easier to endorse by males compared to other items included performing ordinary daily activities (0.67 logits), unhappy about visual condition (0.60 logits), concern about possible worsening of visual condition (0.58 logits), and feeling less inclined to meet people, friends (0.81 logits). The only item that showed large DIF was doing manual activity (1.52 logits) and was rated by females as relatively easy to perform. Two items showed moderate DIF by age: crossing the street (0.63 logits) and reading an article in a newspaper (0.51) both rated by older age group (≥74 years) as relatively easier to perform. One item showed DIF by presence or absence of systemic comorbidity: doing manual activity was rated 1.06 logits easier in the group with systemic comorbidity. Two items showed DIF by the cataract status, i.e., bilateral vs. unilateral cataract: concern with possible worsening of condition (rated 0.50 logits easier in group with bilateral cataract) and problems in perceiving a dip on the ground (rated as 0.55 logits relatively easier by patients awaiting cataract surgery in the fellow eye).

Dimensionality

Principal Components Analysis of the residuals gave the unexplained variance explained by the first contrast as 1.9 eigenvalue units which supports the assumption of unidimensionality.

TABLE 3.
Summary of the Global Fit Statistics for Person Ability and Item Difficulty Parameters for 17-item and 15-item QOL-VFQ

Parameter	Separation index	Reliability	Average infit mean square	Average outfit mean square	Model measurement error	SD
Person ability[a]	2.60	0.87	1.01	0.93	0.60	0.16
Item difficulty[a]	10.64	0.99	1.02	0.94	0.13	0.02
Person ability[b]	2.52	0.86	1.00	0.92	0.64	0.16
Item difficulty[b]	10.80	10.99	1.01	0.93	0.13	0.02

[a] 17 items.
[b] 15 items.

TABLE 4.
Item fit statistics for the QOL-VFQ

Item no.	Item	Item calibration	Standard error square	Infit mean	Outfit mean
12	Read an article	2.37	0.11	1.00	0.99
3	Concern possible worsening	2.24	0.11	1.23	1.46
2	Unhappy about visual condition	1.44	0.11	1.11	1.13
15	Vision in dim light	1.01	0.11	0.95	0.95
9	Reading prices	0.84	0.11	0.82	0.77
14	Vision deteriorate in bright light	0.66	0.12	1.09	1.08
8	Perceiving dip	0.55	0.12	0.85	0.84
13	Doing manual activity	0.12	0.12	1.21	1.05
10	Recognizing people across street	0.11	0.12	1.05	0.97
11	Recognize person in crowded room	−0.68	0.13	0.86	0.74
6	Crossing street	−1.10	0.14	0.96	0.77
1	Ordinary daily activities	−1.50	0.15	0.94	0.70
7	Bump against people	−1.91	0.16	1.02	1.40
5	Feel useless or burden	−1.94	0.16	1.03	0.64
4	Feel less inclined to meet people	−2.20	0.17	0.97	0.51

Criterion Validity

There was a fair and statistically significant relationship ($r = -0.27$, $p < 0.0001$) between the binocular visual acuity and the Rasch scaled QOL-VFQ score.

Performance of Subscales within Rasch Model

We performed a separate Rasch analysis on each subscale as per the approach taken to the full version of the QOL-VFQ above. Table 5 shows the results. The person separation was uniformly below the accepted level (2.0) for each of the subscales indicating their poor discriminatory abilities. Similarly, the targeting was also poor for all the subscales although it was relatively better for the "sensory adaptation" subscale suggesting that the difficulty posed by the items included in this subscale closely match the patient's abilities. However, this subscale also suffered from low reliability indicating that one cannot rely on the person measures obtained. In summary, none of the existing subscales provided a valid measurement and no amount of modification could repair the performance of these subscales. Perhaps, adding more items to these subscales could improve their performance.

DISCUSSION

Reliability Comparison with Other Studies

The original and revised version of the QOL-VFQ both displayed high person separation indices and reliability. This suggests that the QOL-VFQ can differentiate between three strata of patient abilities. The reliability of QOL-VFQ in the current study is comparable with other questionnaires validated using Rasch analysis in cataract patients.[6,19] However, these questionnaires used more response categories in the original version that required collapsing. For example, the ADVS was reported to have a person separation of 2.37 for the disordered 5-point response scale and 2.53 for the revised 3-point response scale. We observed ordered category thresholds for the QOL-VFQ that did not require response categories to be collapsed. This optimal utilization of the categories is perhaps related to the use of only three aptly worded (such as "not at all," "quite a lot," and "very much") categories in the rating scale. These categories suited the higher abilities of our patients.

Targeting

Rasch analysis places the items and persons along the same scale, enabling simultaneous comparison of item difficulty and person ability (Fig. 2). This feature is lacking in the CTT methodology.[8,34] A well targeted instrument has evenly spaced, well-fitting items and the persons are positioned opposite the items at the same level in the person-item map that can provide a good measurement, much like a ruler.[35] However, we observed a large ceiling effect with a small floor effect for the QOL-VFQ because many of our patients had no difficulty, even with the most difficult activity—read newspaper. On visual inspection of the person item map it appears that there are no items that target persons with higher visual ability, but this is not entirely the case because the items use multiple response categories and each item is represented not just by its mean value but levels of endorsement. Such intricate details may not be readily apparent in the person-item map. Nevertheless, the ceiling effect belies the poor targeting from which we can deduce that the instrument would be better suited to a more impaired population. The QOL-VFQ may, therefore, function optimally in a severely visually impaired cataract population, for example, in certain regions such as China and Africa where patients present at a relatively later stage for surgery than in Australia.[36]

The mean person ability for patients awaiting cataract surgery in the second eye was statistically significantly higher (2.72 ± 2.40 vs. 2.21 ± 2.30 logits, Independent Samples test, $F = 0.44$, $p = 0.04$) than for those awaiting surgery in their first eye. This finding is predictable as it would be expected that it would be easier to perform activities after surgery in one eye. This raises the concern of what targeting would be like after second eye cataract surgery; likely it would be even worse. Given the existent problem of a ceiling effect, it is likely that the measurement of surgical outcome with the QOL-VFQ would be impaired by this ceiling effect. The mean person ability (2.44 logits) of entire study sample in the QOL-VFQ suggested that our patients, on an average, had a high level of visual functioning. Although two thirds of our patients were waiting to undergo cataract surgery in their first eye, most of all our patients were, however, not severely visually impaired as evidenced by their mean visual acuity (logMAR 0.22 or $6/9.5^{-1}$). Nevertheless, visual disability is an essential indicator for cataract surgery in our patients. All patients had some disability, although not necessarily for most of the activities listed in the QOL-VFQ questionnaire.

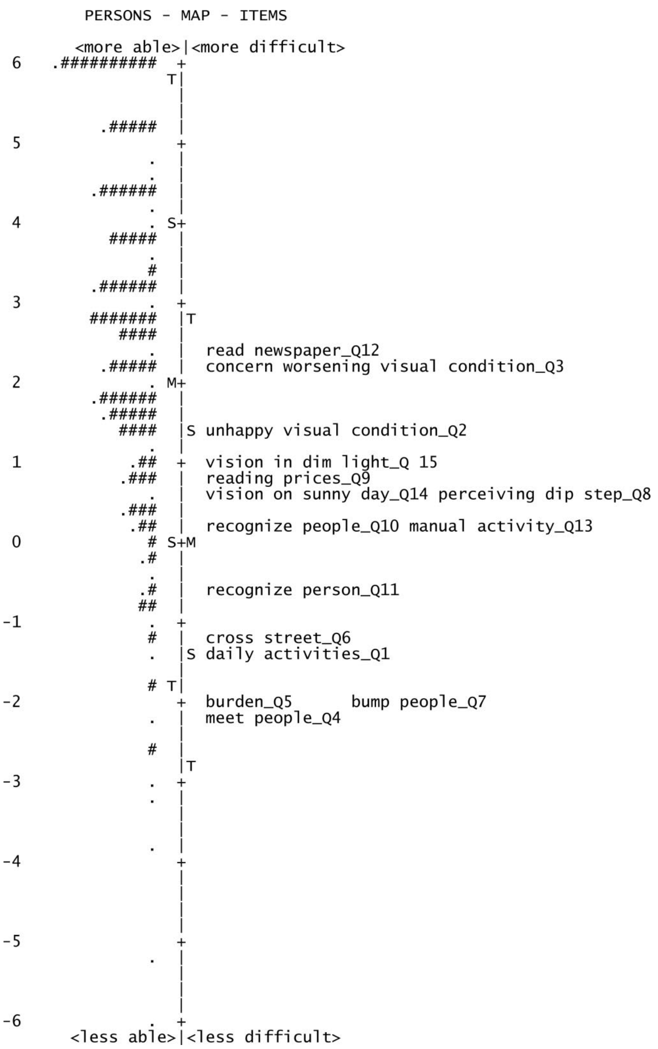

FIGURE 2.

Person-item map for 15-item QOLVFQ. The subjects are on the left of the dashed line and more able subjects are located at the top of the map. Items are located on the right of the dashed line and more difficult items are also located at the top of the map. Each "#" represents four subjects, and each "." represents three subjects. M, mean; S, 1 SD from the mean; T, 2 SD from the mean.

Poor targeting in patients with cataract has been a common feature with other questionnaires too such as the ADVS,[6] Impact of Visual Impairment (IVI),[37] and the 10-item Vision Core Module 1 (VCM1)[38] that have been revalidated previously using Rasch analysis. The VF-14[19] also shares the same shortcoming. Interestingly, all these questionnaires including the QOL-VFQ do not contain enough difficult items relevant to visual disability in patients with cataract. Therefore, similar to suggestions made for the ADVS,[6] VF-14[19] and the Impact of Visual Impairment questionnaire[37] and VCM1,[38] the poor targeting of the QOL-VFQ could be improved by adding more items that match the abilities of people with less visual disability. These items could relate to activities requiring fine resolution such as threading a needle, recognizing faces at a distance, locking/unlocking doors etc. However, adding more items will require further retesting of the instrument's psychometric properties. This problem could be avoided if questionnaires are originally developed using Rasch analysis.[4,5,32,39]

Six items (e.g., recognize person, crossing the street, performing ordinary daily activities, feel burden) were located below the mean of the item group indicating that these were too easy for the patients and do not help in differentiating them. Of these seven items, the least difficult were "meeting people," "feeling burden" and "bump against other people." The item positions for "feeling burden" and "bump against other people" were overlapping by being located along the same line suggestive of possible redundancy.

Item Reduction and Dimensionality

Two items showed misfit. After the removal of these items, the other items showed acceptable fit statistics. The misfitting items were "How much is your driving/vision disturbed by the lights of oncoming cars"? and "How much problem do you have in recognizing colors"? The misfit of the driving item was not surprising as it is not unusual for driving items to misfit a visual disability scale. This suggests that the item is being influenced by something other than visual disability.[6,29,30] The misfit probably occurred because 157 (40%) patients were non-drivers (either had stopped driving due to visual reasons or did not possess a driver's license), whereas the remaining patients had higher levels of difficulty with this

TABLE 5.
Results of testing of subscale fit to the Rasch model for QOL-VFQ

	Subscale				
Parameter	Overall assessment of visual satisfaction	Self-assessment of visual field	Self-assessment of distance visual acuity	Self-assessment of near visual acuity	Self-assessment of color vision
Person separation	1.35	0.70	0.99	0	0.45
Person separation reliability	0.65	0.33	0.50	0	0.17
Mean item location	0	0	0	0	0
Mean person location	3.63	4.87	3.09	1.81	1.51

activity (i.e., they chose the higher category). Removing this item did not significantly influence the separation indices. This finding lends credence to the notion that this item was producing as much noise as signal, because removal of a good fitting item should ideally cause a loss of precision.[35]

Rasch analysis demonstrated the 15-item QOL-VFQ to be unidimensional suggesting that all the items were indeed measuring a single underlying trait, i.e., visual disability. This has clinical implications in that a single summary score can now be used to represent the visual disability arising from cataract that could help making decisions in prioritizations for surgery; although the poor targeting limits its value for these applications. Our finding of unidimensionality for the QOL-VFQ is consistent with the VCM1,[38] another questionnaire that was revalidated using Rasch analysis in a cataract population.

Differential Item Functioning

As in the current study, it is not uncommon to come across DIF with a large sample. The power to detect DIF increases with sample size.[40] There was, however, some large DIF for gender related to performing manual activity which women seemed to consider represented easier tasks than men did. The level of gender bias found with this item suggests its deletion, which we consider a valid option. However, deleting items should be considered cautiously as deleting a good fitting item could lead to a lowering of its reliability and validity.[40] The remaining items that showed DIF were probably at too low a level to cause great concern. Moreover, of all the items that showed DIF, six showed DIF in one direction whereas the others showed DIF in the opposite direction. This suggests that the DIF for the QOL-VFQ is of limited consequence for its measurement properties. However, DIF was entirely absent in our previously Rasch revalidated questionnaires, the IVI,[37] and VCM1[38] in a similar cataract population.

Performance of Subscales

None of the five subscales functioned well. All attempts to repair functioning of the subscales, for example removing less well fitting items, did not help overcome these problems. The fundamental problem is that all of these subscales have too few items to discriminate the patient population; the addition of items to each subscale is required. This finding is consistent with previous studies that found for subscales to function satisfactorily, sufficient length was a key attribute.[41]

Although it is ideal that investigators wishing to use the QOL-VFQ obtain Rasch scores for their own data by performing Rasch analysis, it may not always be realistic and yet they may want to utilize the full potential of Rasch analysis. Furthermore, if future applications of the QOL-VFQ in another population believe that their sample is significantly different from that of the present study (i.e., an Australian population), then we suggest that researchers perform Rasch analysis for their own sample. However, as indicated as one of aims we developed an Excel sheet (available from the authors or from the journal's website) for ready conversion of raw scores to Rasch-scaled scores (in logits). All these steps will encourage clinicians and researchers to use patient-reported outcomes more frequently in their evaluation of outcomes of cataract surgery.

The provision of spreadsheet for the conversion of raw responses to Rasch measurement enables other researchers to use this questionnaire and gain the benefits of interval measurement. However, there are other questionnaires that could equally be used. Any existing instrument can be scored with Rasch analysis, but without thorough revalidation, problems like item misfit may exist. Several questionnaires have been revalidated using Rasch analysis and some of these have conversion spreadsheets provided—IVI,[37] VCM1.[38] A better approach would be to develop new questionnaires using Rasch analysis that would provide better targeting in particular. However, the next step really should be the development of an item bank and computer-adaptive testing in which a large number of items can be tailored to provide the maximum amount of information based on a participant's response to previous items. Such item banks exist for other areas of health care,[42–44] but a similar approach for ophthalmology is lacking which calls for the development of such item banks.

CONCLUSIONS

In conclusion, we found that QOL-VFQ could be improved by slightly modifying the original QOL-VFQ items and by shortening it to 15 items. The modified QOL-VFQ performed well on most aspects of the assessment: it has good discriminative ability, unidimensionality, items all fit the construct, the response scale functions well and there is a small amount of DIF which is probably not of a magnitude to be problematic. The only serious problem for the overall scale was the poor targeting of item difficulty to person ability in this cataract population from Australia. Due to the poor performance of the subscales we do not recommend that these subscales be used.

ACKNOWLEDGMENTS

We thank all our patients who volunteered to participate in the study.

This work was supported in part by National Health and Medical Research Council (Canberra, Australia) Centre of Clinical Research Excellence Grant 264620.

Konrad Pesudovs is supported by National Health and Medical Research Council (Canberra, Australia) Career Development Award 426765.

The authors have no personal financial interest in the development, production, or sale of any device discussed herein.

Received November 17, 2008; accepted January 22, 2009.

REFERENCES

1. Steinberg EP, Tielsch JM, Schein OD, Javitt JC, Sharkey P, Cassard SD, Legro MW, Diener-West M, Bass EB, Damiano AM, Steinwachs DM, Sommer A. The VF-14. An index of functional impairment in patients with cataract. Arch Ophthalmol 1994;112:630–8.
2. Mangione CM, Phillips RS, Seddon JM, Lawrence MG, Cook EF, Dailey R, Goldman L. Development of the 'Activities of Daily Vision Scale'. A measure of visual functional status. Med Care 1992;30: 1111–26.
3. Pesudovs K, Elliott DB, Coster DJ. The Cataract Outcomes Questionnaire—a Rasch scaled measure of visual disability. Invest Ophthalmol Vis Sci 2005;46:E–abstract 3844.
4. Pesudovs K, Garamendi E, Elliott DB. The Quality of Life Impact of

Refractive Correction (QIRC) Questionnaire: development and validation. Optom Vis Sci 2004;81:769–77.
5. Pesudovs K, Garamendi E, Elliott DB. The Contact Lens Impact on Quality of Life (CLIQ) Questionnaire: development and validation. Invest Ophthalmol Vis Sci 2006;47:2789–96.
6. 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;44:2892–9.
7. Massof RW. The measurement of vision disability. Optom Vis Sci 2002;79:516–52.
8. DeVellis RF. Classical test theory. Med Care 2006;44:S50–9.
9. Crocker L, Algina J. Introduction to Classical and Modern Test Theory. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers; 1986.
10. Nunnally JC, Bernstein IH. Psychometric Theory, 3rd ed. New York, NY: McGraw-Hill; 1994.
11. Carta A, Braccio L, Belpoliti M, Soliani L, Sartore F, Gandolfi SA, Maraini G. Self-assessment of the quality of vision: association of questionnaire score with objective clinical tests. Curr Eye Res 1998; 17:506–11.
12. 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:I25–36.
13. White LJ, Velozo CA. The use of Rasch measurement to improve the Oswestry classification scheme. Arch Phys Med Rehabil 2002;83: 822–31.
14. Raczek AE, Ware JE, Bjorner JB, Gandek B, Haley SM, Aaronson NK, Apolone G, Bech P, Brazier JE, Bullinger M, Sullivan M. Comparison of Rasch and summated rating scales constructed from SF-36 physical functioning items in seven countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998;51:1203–14.
15. Mallinson T. Why measurement matters for measuring patient vision outcomes. Optom Vis Sci 2007;84:675–82.
16. Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of Item Response Theory. Newbury Park, CA: Sage Publications; 1991.
17. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to their Development and Use, 2nd ed. Oxford, UK: Oxford University Press; 1995.
18. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago, IL: University of Chicago Press; 1980.
19. 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;4:667–80.
20. Kirkwood BJ, Pesudovs K, Latimer P, Coster DJ. The efficacy of a nurse-led preoperative cataract assessment and postoperative care clinic. Med J Aust 2006;184:278–81.
21. Elliott DB, Hurst MA, Weatherill J. Comparing clinical tests of visual function in cataract with the patient's perceived visual disability. Eye 1990;4:712–7.
22. Rubin GS, Bandeen-Roche K, Huang GH, Munoz B, Schein OD, Fried LP, West SK. The association of multiple visual impairments with self-reported visual disability: SEE project. Invest Ophthalmol Vis Sci 2001;42:64–72.
23. Wright BD, Stone MH. Best Test Design. Chicago, IL: MESA Press; 1979.
24. Andrich D. A model for contingency tables having an ordered response classification. Biometrics 1979;35:403–15.
25. Linacre JM. WINSTEPS Rasch Measurement [Computer Program]. Chicago, IL: Winsteps.com; 2008.
26. Wolfe F, Michaud K, Pincus T. Development and validation of the health assessment questionnaire II: a revised version of the health assessment questionnaire. Arthritis Rheum 2004;50:3296–305.
27. Comins J, Brodersen J, Krogsgaard M, Beyer N. Rasch analysis of the Knee injury and Osteoarthritis Outcome Score (KOOS): a statistical re-evaluation. Scand J Med Sci Sports 2008;18:336–45.
28. Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Mahweh, NJ: L. Erlbaum; 2001.
29. Smith RM. Person fit in the Rasch model. Educ Psychol Meas 1986; 46:359–72.
30. 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.
31. Conrad KJ, Smith EV Jr. International conference on objective measurement: applications of Rasch analysis in health care. Med Care 2004;42:I1–6.
32. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. Optom Vis Sci 2007;84: 663–74.
33. Angoff WH. Use of difficulty and discrimination indices for detecting item bias. In: Ber RA, ed. Handbook of Methods for Detecting Test Bias. Baltimore, MD: Johns Hopkins University Press;1982: 96–116.
34. Crocker L, Algina J. Introduction to Classical and Modern Test Theory. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers; 1986.
35. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. Med Care 2004;42:I17–24.
36. Foster A. Vision 2020: the cataract challenge. Community Eye Health 2000;13:17–9.
37. Pesudovs K, Caudle LE, Rees G, Lamoureux EL. Validity of a visual impairment questionnaire in measuring cataract surgery outcomes. J Cataract Refract Surg 2008;34:925–33.
38. Lamoureux EL, Pesudovs K, Pallant JF, Rees G, Hassell JB, Caudle LE, Keeffe JE. An evaluation of the 10-item vision core measure 1 (VCM1) scale (the Core Module of the Vision-Related Quality of Life scale) using Rasch analysis. Ophthalmic Epidemiol 2008;15: 224–33.
39. du Toit R, Palagyi A, Ramke J, Brian G, Lamoureux EL. Development and validation of a vision-specific quality-of-life questionnaire for Timor-Leste. Invest Ophthalmol Vis Sci 2008;49:4284–9.
40. Hambleton RK. Good practices for identifying differential item functioning. Med Care 2006;44:S182–8.
41. Lamoureux EL, Pallant JF, Pesudovs K, Rees G, Hassell JB, Keeffe JE. The impact of vision impairment questionnaire: an assessment of its domain structure using confirmatory factor analysis and Rasch analysis. Invest Ophthalmol Vis Sci 2007;48:1001–6.
42. Walter OB, Becker J, Bjorner JB, Fliege H, Klapp BF, Rose M. Development and evaluation of a computer adaptive test for 'Anxiety' (Anxiety-CAT). Qual Life Res 2007;16(Suppl 1):143–55.
43. Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). Qual Life Res 2005;14:2277–91.
44. Lai JS, Cella D, Dineen K, Bode R, Von Roenn J, Gershon RC, Shevrin D. An item bank was created to improve the measurement of cancer-related fatigue. J Clin Epidemiol 2005;58:190–7.

Konrad Pesudovs
Department of Ophthalmology
NH&MRC Centre for Clinical Eye Research
Flinders Medical Centre, Bedford Park
South Australia 5042, Australia
e-mail: Konrad.Pesudovs@flinders.edu.au