Low Vision

Evaluation of a Vision-Related Utility Instrument: The German Vision and Quality of Life Index

Robert P. Finger,[1,2] *Karsten Kortuem,*[3] *Eva Fenwick,*[1] *Bettina von Livonius,*[3] *Jill E. Keeffe,*[1] *and Christoph W. Hirneiss*[3]

Purpose. Multi-attribute utility instruments (MAUIs), which contain a descriptive system, including several health dimensions with associated levels of increasing severity, are used commonly to measure utilities. However, the validity of the descriptive systems rarely is examined using modern psychometric theory. Therefore, we evaluated the psychometric properties of the German version of the Vision and Quality of Life Index (VisQol), a six-item vision-related MAUI.

Methods. The German VisQol was self-administered to 340 patients and 280 controls. All subjects underwent a full ophthalmologic examination, including best-corrected visual acuity (VA) testing. The psychometric properties of the VisQoL were assessed using Rasch analysis.

Results. The VisQoL's descriptive system did not function in controls. In patients, after collapsing response categories to resolve disordered thresholds and omitting misfitting persons, the measurement properties (i.e., precision, unidimensionality, and targeting) of the German VisQoL were satisfactory. Most person misfit related to unexpected responses to item 4 ("organizing assistance"). Rasch-generated person estimates were not different between age categories, sex, or underlying ocular condition, but decreased significantly with presence of visual impairment in the better eye (LogMAR $\geq$ 0.5, 1.20 $\pm$ 4.62 compared to 3.46 $\pm$ 3.52, $P < 0.001$).

Conclusions. The German VisQoL's descriptive system displayed adequate fit to the Rasch model after removal of a large proportion of patients with poor fit statistics. However, the wording of item four should be revised to reduce respondent confusion and measurement "noise." The scale's descriptive system does not function in a sample of visually unimpaired persons, most likely due to a lack of variance in the measured trait. (*Invest Ophthalmol Vis Sci.* 2013;54:1289–1294) DOI: 10.1167/iovs.12-10828

From the [1]Centre for Eye Research Australia, University of Melbourne, Royal Victorian Eye and Ear Hospital, Melbourne, Australia; the [2]Department of Ophthalmology, University of Bonn, Bonn, Germany; and the [3]Department of Ophthalmology, Ludwig-Maximilians University, Munich, Germany.

Supported by grants from Novartis Pharma Germany (CWH) and the German research Council (Grant DFG FI 1540/5-1, RPF). CERA receives Operational Infrastructure Support from the Victorian Government. The authors alone are responsible for the content and writing of this paper.

Submitted for publication August 22, 2012; revised September 29 and November 18, 2012, and January 4, 2013; accepted January 8, 2013.

Disclosure: **R.P. Finger**, None; **K. Kortuem**, None; **E. Fenwick**, None; **B. von Livonius**, None; **J.E. Keeffe**, None; **C.W. Hirneiss**, None

Corresponding author: Christoph W. Hirneiss, Augenklinik der Ludwig-Maximilians-Universität, Mathildenstr. 8, D-80336 München, Germany; Christoph.Hirneiss@med.uni-muenchen.de.

As demands on healthcare continue to increase, economic evaluations of disease, impairment and interventions also have become increasingly important.[1] A fundamental aspect of every economic evaluation is the collection or inference of utilities that then are used to calculate Quality Adjusted Life Years (QALYs) and, subsequently, the respective impact of disease, impairment, and interventions.[1]

One way of calculating utilities is through multi-attribute utility instruments (MAUIs) in which values are elicited indirectly through patient ratings of their health status from a multifeatured classification system, which captures health-related quality of life (QoL) and allows the comparison across different health states. The only vision-specific MAUI available currently is the 6-item Vision and Quality of Life Index (VisQoL), which was developed and validated specifically for vision-impaired populations.[2] To date, the VisQoL is available only in English, and has not been evaluated using modern psychometric theory, such as Rasch analysis, a form of item response theory. The VisQoL has been shown to perform well using classic test theory and its items were chosen based on item response theory different from the Rasch model.[2,3] However, as the scale provides ordinal measurement of vision-related quality of life (VRQoL) based on its Likert scored descriptive system, and is summarized into a single score following conversion, Rasch analysis is useful not only in assessing the VisQoL's measurement properties, but also in transforming the ordinal score into an interval-level linear estimate.[4]

Therefore, in our study, we determined the validity, reliability and measurement characteristics of the German VisQoL's descriptive system using Rasch analysis, and investigated the relationship between the severity of vision impairment, main causes of vision loss, and VisQoL measurement in a German sample of patients with and without vision impairment as well as healthy controls.

Patients and Methods

Patients and controls were recruited from the outpatient clinic at the department of ophthalmology, University of Munich between August 2011 and June 2012. Institutional review board (IRB) approval was obtained from the IRB of the University of Munich. All patients gave signed informed consent for study participation before enrollment. The study adhered to the tenets of the Declaration of Helsinki.

Participants

Participants underwent a complete ophthalmic examination, including presenting and best-corrected visual acuity (VA), biomicroscopy, intraocular pressure measurements, and funduscopy. Further diagnostic tests (visual field assessment, fluorescein angiography, optical coherence tomography, and electrophysiology) were performed as appropriate in each individual case. VA was tested as best-corrected

TABLE 1. The VisQoL Items, Original Response Categories and Collapse of Response Categories following Rasch Analysis

Item	Response Categories	Collapse of Response Categories
Does my vision make it likely I will injure myself?	5: "Most unlikely" – "almost certainly"	12345 → 12223
Does my vision make it difficult to cope with the demands in my life?	6: "No effect" – "unable to"	123456 → 122223
Does my vision affect my ability to have friendships?	7: "No effect" – "unable to" + not applicable	1234567 → 123334; 7 = missing
Do I have difficulty organizing any assistance I may need?	6: "No difficulty" – "unable to" + not applicable	123456 → 12223; 6 = missing
Does my vision make it difficult to fulfill the roles I would like to fulfill in life?	6: "No effect" – "unable to"	123456 → 123334
Does my vision affect my confidence to join everyday activities?	6: "More confident" – "not at all"	123456 → 123334

distance VA using a standard retro-illuminated LogMar chart at 4 meters. Ability of participants to complete the questionnaires was assessed by interviewers before recruitment. The questionnaires were self-administered.

The VisQoL Index

The VisQoL comprises a descriptive system that covers six dimensions of self-reported VRQoL: physical well-being, independence, social well-being, self-actualization, and planning and organization, with every dimension being represented by one item (Table 1).[2,3] Each question was preceded by "Does my vision…" and each dimension had between five and seven response categories, ranging from, for example, "no effect" to "unable to do." Two dimensions also have a "nonapplicable" option. Response categories vary from five (item 1), to six (item 2, 4, 5, 6) to seven (item 3). The VisQoL was developed originally as part of the Assessment of Quality of Life (AQoL) instrument 7-D, as its seventh dimension and can be used in conjunction with the AQoL or by itself.[5]

The VisQoL measures VRQoL, which then can be translated into health states defined by the VisQoL responses, using an available value set derived from participant surveys using the Time Trade Off (TTO) method yielding vision-related utilities.[2] Utilities were not generated in this study, which is concerned solely with assessing the psychometric properties of the descriptive system of the VisQoL.

German Language Translation of the VisQoL

The VisQoL was translated into German and back translated into English, and subsequently a final version created based on this process. The final translation then was pilot tested in focus groups of patients, discussing wording, comprehension, and cultural appropriateness of content. This resulted in several changes of the wording of questions, but no change of the overall content was necessary.

Psychometric Evaluation of the German VisQoL

Rasch analysis is a psychometric method that describes mathematically the interaction between respondents and test items, and applies a strict model that the pattern of participants' responses should satisfy.[6–9] During Rasch analysis, scores that approximate interval-level measurement (expressed in log of the odds units, or logits) are estimated from raw ordinal data. Rasch analysis also provides greater insight into the psychometric properties of the instrument compared to traditional methods. Several techniques are available to determine how well items fit the latent trait being measured, how well the items discriminate between the respondents, and how well item difficulty targets person ability.[10] We used the following criteria to assess the psychometric properties of the VisQoL.

Threshold Ordering. To determine whether the categories used to rate the VisQoL items are valid, we assessed the response category threshold ordering. First, over- or underutilization of response categories, as well as ability of respondents to discriminate between response categories, was assessed. Disordered thresholds, if evident, were addressed by collapsing categories.

Precision of the Instrument. The ability of the scale to discriminate between different levels of person ability was assessed using person separation index (PSI) and person reliability (PR) scores. Values of >2.0 and >0.8, respectively, were considered adequate and represented the capacity of the scale to distinguish three levels of person ability.

Unidimensionality. Whether the scale measured a single latent trait was assessed in two ways. First, we tested how well each item "fits" or "misfits" the underlying trait through an "infit" mean square standardized residuals (MNSQ) statistic.[11] A value of 0.7 to 1.3 was considered acceptable, while lower or higher values may indicate redundancy or unacceptable variation in the responses, respectively. Second, the items were tested for local independence using Principal Components Analysis (PCA), which means that they were not related except for the fact that they measured the same trait, with as little overlap between items as possible. The PCA of residuals for the first factor should exceed 50% and the first contrast of residuals should be <2.0 eigenvalues.[11]

Targeting. The targeting of the instrument was determined by visual inspection of the person-item map, and calculation of the difference between item and person means. A difference of >1.0 logits suggested that the difficulty of the items does not target the ability of the sample participants adequately.

Person "fit." The extent to which the responses of any person conformed to the Rasch model expectation were determined using the infit mean square (MnSq) fit statistic**.** Person misfit indicated more erratic or haphazard performance than predicted by the Rasch model.[12] In our study, any person with an infit MnSq score >2 was removed from the analysis as this level of misfit distorted the measurement system. Removal of misfitting persons generally improved other measurement characteristics, such as scale precision.

Differential Item Functioning (DIF). Each item was assessed for DIF, which is a statistical method for detecting whether sample subgroups (e.g., sex, age groups) systematically respond differently to certain items, despite having similar underlying ability. A DIF contrast of >1.0 logits for an item indicated that interpretation of the item may be biased for some participant subgroups.

We performed Rasch analysis on the German VisQoL using Winsteps software (version 3.68; Winsteps, Chicago, IL). The Andrich rating scale model was used for the VisQoL.[12]

Statistical Analyses

The SPSS statistical software (Version 19.0; SPSS Science, Chicago, IL) was used to analyze the data. Descriptive statistical analyses were performed to characterize the participants' sociodemographic, clinical, and VisQoL data. VA was categorized into two categories: Normal

TABLE 2. Characteristics of the Sample, Patients Only

	Total Sample (n = 340)	Sample after Removing Misfitting Persons (n = 257)	VisQol Rasch-Derived Person Measures (n = 257, in Logits)	
	n (%) or Mean ± SD	*n* (%) or Mean ± SD	Mean ± SD	*P**
Age	66.25 ± 14.47	66.80 ± 14.58		
Better eye VA in LogMAR	0.28 ± 0.37	0.29 ± 0.38		
Total score			2.95 ± 3.90	
Age categories				
<65	132 (38.9%)	92 (35.9%)	3.05 ± 3.68	0.737
65+	207 (61.1%)	164 (64.1%)	2.88 ± 4.03	
Sex				
Male	151 (44.8%)	116 (45.5%)	2.73 ± 3.96	0.459
Female	186 (55.2%)	139 (54.5%)	3.09 ± 3.85	
VI ≥0.5 LogMAR				
Yes	68 (20.2%)	55 (21.7%)	1.20 ± 4.62	<0.001
No	269 (79.8%)	199 (78.3%)	3.46 ± 3.52	
Ocular condition (patients only)				
AMD	126 (37.4%)	102 (40.0%)	2.37 ± 4.13	0.065
DED	107 (31.8%)	80 (31.4%)	3.90 ± 3.54	
RVO	15 (4.5%)	10 (3.9%)	1.53 ± 2.71	
Glaucoma	22 (6.5%)	13 (5.1%)	2.96 ± 4.26	
Cataract	16 (4.7%)	11 (4.3%)	4.21 ± 2.93	
Other	51 (15.1%)	39 (15.3%)	2.37 ± 4.05	

LogMAR, logarithm of the minimum angle of resolution; RVO, retinal vascular occlusion; VI, vision impairment.
* One-way ANOVA.

vision or mild-to-moderate vision impairment in the better eye (<0.5 logarithm of the minimum angle of resolution [–LogMAR]), and moderate-to-severe vision impairment and blindness (–LogMAR ≥0.5). All tests were considered to be statistically significant at an adjusted level of $P < 0.05$.

RESULTS

Sample Characteristics

A total of 340 patients was included in this study. The mean age of the sample was 66 ± 15 years (± SD). Just over half of the sample (55%) were female, and 20% were visually impaired with a VA ≥0.5 LogMAR (Table 2). Most patients (37%) suffered from age-related macular degeneration (AMD), followed by diabetic eye disease (DED, including diabetic retinopathy and maculopathy) in 32% and glaucoma in 7% (Table 2). The 280 healthy controls had a mean age of 46 ± 9 years and a better eye VA of 0.04 ± 0.09 LogMAR, and 72% were female and no one was visually impaired (data not shown).

Psychometric Evaluation of the VisQoL

Based on the varying number of response categories, and the different wording of the items, three rating scales were applied. As the suggested coding of response categories leads to higher ability being assigned a lower numerical value, the rating scales were reversed for analysis, with higher scores indicating higher ability, that is, less visual impairment. Category thresholds were disordered for all rating scales, indicating an inability by participants to differentiate sufficiently between them. Accordingly, response categories were collapsed to only three (item 1, 2, 4, 5) and four (item 3 & 6) response categories (Table 1, example presented in Fig. 1).

In our overall sample of patients and controls, a number of items displayed misfit, and all other measurement characteristics displayed suboptimal fit to the Rasch model (Table 3). As our patient sample represents a broad spectrum of visual ability, we excluded the controls from the further analysis as the inclusion of a large group of healthy controls, that is, unimpaired persons with no or little understanding of the impairment or underlying trait measured, is likely to obscure measurement.

Further psychometric assessment of the VisQoL in patients demonstrated poor precision, one misfitting item (item 4, "organizing assistance"), and DIF for 4 items (Table 3). After removing item 4, overall fit statistics did not improve, person separation and reliability remained suboptimal, and the targeting and DIF worsened. Thus, item 4 was retained. Examination of person fit statistics revealed a number of severely misfitting participants (infit MnSq >2, n = 83). Misfitting persons were not significantly different from non-misfitting persons in regards to their age, sex, underlying ocular conditions, better eye VA or visual impairment (Table 2). Rather, most misfit seemed to be due to unexpected answers to item 4, with a large proportion of participants answering either the extreme ("not able") or "not applicable." Upon removal of misfitting persons, all other fit statistics improved to an acceptable level. For example, no items displayed misfit or DIF, and the PSI of 2.1 demonstrated the ability of the VisQoL to discriminate between at least three different levels of person ability. Targeting of the scale remained suboptimal, and the person item map (Fig. 2) demonstrated an evident lack of items targeting the more able participants. Taken together, these fit parameters indicated that the German VisQoL is a valid and reliable scale in our patient sample.

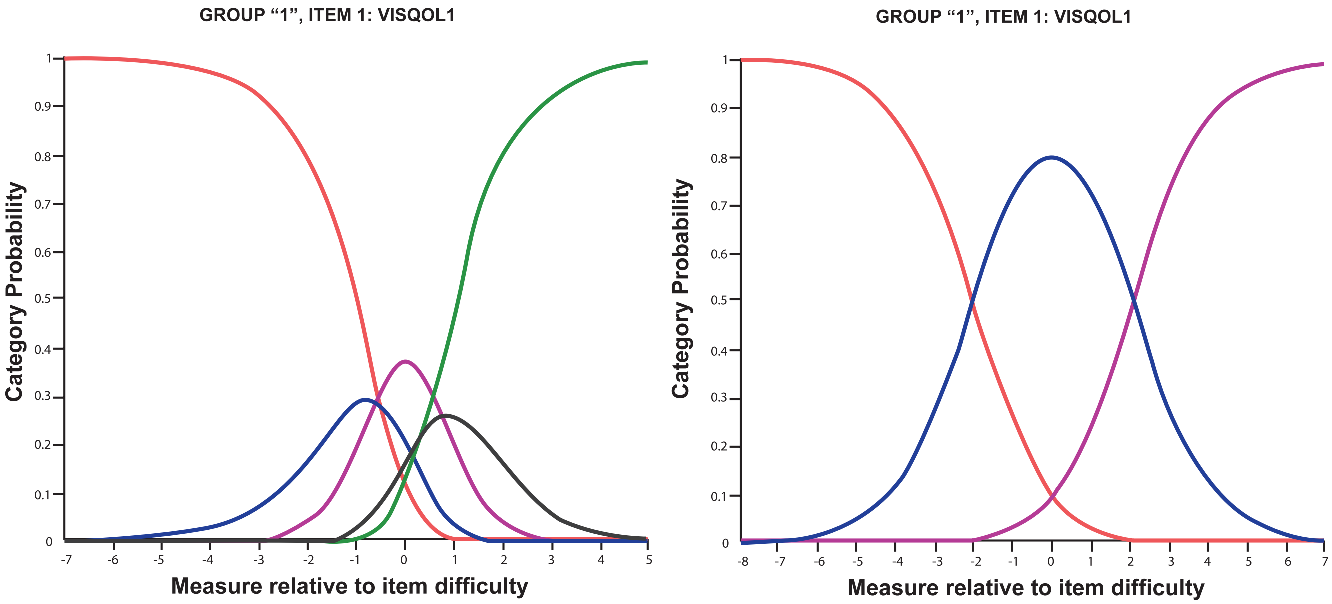

Figure 1. Response categories for rating scale 1 showing disordered thresholds before (*left*) and ordered thresholds after (*right*) collapsing of responses from 5 to 3.

Factors Associated with VisQoL Measurement

The VisQoL person measures were not associated with age, sex, or underlying ocular condition, but decreased significantly with presence of visual impairment (LogMAR $\geq$0.5, 1.20 $\pm$ 4.62 compared to 3.46 $\pm$ 3.52, $P < 0.001$, Table 2).

Discussion

The German VisQoL's descriptive system displayed adequate fit to the Rasch model in our sample of patients with and without vision impairment, after collapsing response categories and removal of a considerable number of misfitting persons. Most person misfit was due to unexpected answers to item 4 ("organizing assistance") which must be rephrased to reduce the amount of noise created in the overall measurement. Following these revisions, the German VisQoL provided satisfactory measurement of VRQoL in patients in our sample, and is responsive to the presence of at least moderate visual impairment (LogMAR $\geq$0.5) irrespective of the underlying ocular condition. The scale does not function in visually unimpaired, healthy controls.

The VisQoL demonstrated suboptimal targeting of person ability to item difficulty. Item difficulty is determined by item content, but also by wording and response categories.[13] Intuitively, all VisQoL items are quite "easy" (e.g., "friendships," "fulfill roles," "confidence"), which explains why they

Table 3. The Fit Parameters of the German VisQoL Compared to the Rasch Model

Parameters	Rasch Model	Total Sample, Patients n = 340; Controls n = 280	VisQol Patients Only, n = 340	VisQol Misfitting Patients Removed, n = 257
Item no.		1–6	1–6	1–6
N of misfitting items	0	infit MnSq No. 4: 2.21 No. 2: 0.60 No. 3: 0.64 No. 5: 0.58 outfit MnSq No. 4: 2.04 No. 2: 1.23 No. 3: 0.58 No. 5: 0.56	infit MnSq No. 4: 2.07 outfit MnSq No. 4:1.95	0
Person separation (PSI)	>2.0	1.34	1.39	2.10
Person reliability (PR)	>0.8	0.64	0.66	0.81
Person mean	0	-1.75	1.32	2.95
PCA; Eigenvalue for first contrast	<2.0	1.8	1.5	1.5
Variance by the first factor	>50%	43.9%	45.0%	65.6%
DIF(contrast)				
Age	<1.0	Not tested	Item 1, 4, 6	None
Sex	<1.0	Not tested	Item 6	None
Vision impairment	<1.0	Not tested	None	None

```
             PERSON - MAP - ITEM
                 <more>|<rare>
 15                  .  +
                        |
 14                     +
                        |
 13                     +
                     .  |
 12                     +
                        |
 11                     +
                       T|
 10                     +
                        |
  9                     +
          ########      |
  8                  .  +
                        |
  7                     +
                       S|
  6                     +
           .######      |
  5                     +
         .########      |
  4                     +T VISQOL6
        ##########      |
  3                  . M+
        .#########      |
  2                     +S
                        |
  1                     +
            .#####      |  VISQOL2
  0                     +M
                        |  VISQOL3  VISQOL5
 -1                  . S+
                        |
 -2                     +S VISQOL1  VISQOL4
      .##########       |
 -3                     +
                        |
 -4                     +T
                 ##    T|
 -5                     +
                  #     |
 -6                     +
                  .     |
 -7                     +
                 <less>|<frequ>
EACH "#" IS 4. EACH "." IS 1 TO 3
```

Figure 2. Person-Item map for the 6 VisQoL items (*right side*) and the final sample of patients in which misfitting persons (infit MnSq >2.0) were removed (*left side*). The *map* indicates a lack of item coverage at the extreme ends, which is more pronounced in nonvisually impaired participants (anyone >0).

may not apply to a visually unimpaired sample. Similar targeting problems have been found for a large number of vision-related functioning instruments, with no or very few items targeting the most able participants.[13] This was thought to be due to most instruments having been developed with visually impaired persons, thus containing content most appropriate to them. The targeting of the VisQoL may be improved by the addition of more "difficult" items.

Item 4 ("Do I have difficulty organizing any assistance I may need"?) created considerable misfit among persons, which may be due to its ambiguous phrasing, as it remains unclear whether the emphasis is on organizing, the need for assistance, or the level of assistance needed. Future studies should assess whether improvements to the phrasing of item 4, for example, "How often do I need assistance"? reduces the misfit associated with this item. Removal of this item did not improve the Rasch statistics, but removal of misfitting persons, whose misfit was due mostly to item 4, worsened the overall targeting of the instrument. As the proportion of participants who displayed misfit was comparably large, the German VisQoL must be revised to resolve the issues with item 4.

While our study has shown that the VisQoL's descriptive system is valid and reliable in a patient sample, it did not function optimally in a "control" group, that is, those without ocular disease or vision impairment, due to an expected lack of variance in the sample and the observed lack of more difficult items targeting nonvisually impaired participants. The VisQoL's descriptive system was never developed to be used in normals (normally-sighted controls) and, as with other scales measuring an impairment or its impact, the inclusion of a large group of healthy controls, that is, unimpaired persons with no or little understanding of the impairment/underlying trait measured, is likely to obscure measurement. This may have implications for generating utilities from the VisQoL descriptive system, given that utilities are required in patient and control samples for economic assessments. All this raises an interesting theoretical question about the use of MAUI descriptive systems—which must conform to the Rasch or another IRT model to ensure valid measurement of the underlying trait—to generate utilities. Thus, any utility conversion based on MAUIs likely will need to be validated using time trade-off or standard gamble interviews with patients as well as controls to allow for a valid conversion of VRQoL scores to arrive at societal preferences.

Strengths of our study included the relatively large sample size and the use of Rasch analysis to assess psychometric properties of the VisQoL, and to produce interval-level measurements of vision-related utilities. Such a psychometric evaluation has not been done for the VisQoL to date to our knowledge. Our study setting in daily clinical routine better reflects actual participant context and their preferences than highly standardized clinical trials. To obtain adequate fit to the Rasch model, a large number of misfitting persons required removal so as to reduce measurement noise. Although removal of these persons reduced the sample size and potentially could contribute to sample bias, the misfitting persons did not differ according to age, sex, better eye VA, or underlying ocular conditions compared to the rest of the sample. This supports our conclusion that the VisQoL's descriptive system provides valid measurement of VRQoL in this sample of patients with and without vision impairment. Another limitation of our study was the lack of concurrent use of other MAUIs with vision content, which could have been used for cross validation.

A number of different IRT models other than Rasch analysis, could have been used in our study. However, Rasch analysis is considered a gold standard assessment in the field of low vision as it applies a strict model to which the data must conform, which is most appropriate in a validation study. Therefore, we did not use other IRT models in our study.

In conclusion, the German VisQoL's descriptive system is a valid instrument to assess VRQoL in a sample of eye patients with and without vision impairment after a series of minor revisions guided by Rasch analysis, as well as removal of a number of misfitting persons. Future studies should assess whether rephrasing of item 4 reduces person and item misfit, and whether the addition of more "difficult" items improves scale targeting, and also whether the VisQoL allows for the generating of utilities in a sample with patients and controls based on its descriptive system.

Acknowledgments

We thank all participants who volunteered their time in contributing.

References

1. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford: Oxford University Press; 2005.
2. Peacock S, Misajon R, Iezzi A, Richardson J, Hawthorne G, Keeffe J. Vision and quality of life: development of methods for the VisQoL vision-related utility instrument. *Ophthalmic Epidemiol*. 2008;15:218–223.
3. Misajon R, Hawthorne G, Richardson J, et al. Vision and quality of life: the development of a utility measure. *Invest Ophthalmol Vis Sci*. 2005;46:4007–4015.
4. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum*. 2007;57:1358–1362.
5. Richardson J, Iezzi A, Peacock S, et al. Utility weights for the vision-related Assessment of Quality of Life (AQoL)-7D instrument. *Ophthalmic Epidemiol*. 2012;19:172–182.
6. 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.
7. 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–I136.
8. Pesudovs K. Patient-centred measurement in ophthalmology - a paradigm shift. *BMC Ophthalmol*. 2006;6:25.
9. Pesudovs K. Autorefraction as an outcome measure of laser in situ keratomileusis. *J Cataract Refract Surg*. 2004;30:1921–1928.
10. Lamoureux E, Pesudovs K. Vision-specific quality-of-life research: a need to improve the quality. *Am J Ophthalmol*. 2011;151:195–197.e2.
11. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci*. 2007;84:663–674.
12. Linacre JM. *A User's Guide to Winsteps/Ministeps Rasch-Model Programs*. Chicago, IL: MESA Press; 2005.
13. Khadka J, McAlinden C, Gothwal VK, Lamoureux EL, Pesudovs K. The importance of rating scale design in the measurement of patient-reported outcomes using questionnaires or item banks. *Invest Ophthalmol Vis Sci*. 2012;53:4042–4054.