The Development of an Instrument to Measure Quality of Vision: The Quality of Vision (QoV) Questionnaire

Colm McAlinden,[1,2] *Konrad Pesudovs,*[3] *and Jonathan E. Moore*[2,4]

Purpose. To develop an instrument to measure subjective quality of vision: the Quality of Vision (QoV) questionnaire.

Methods. A 30-item instrument was designed with 10 symptoms rated in each of three scales (frequency, severity, and bothersome). The QoV was completed by 900 subjects in groups of spectacle wearers, contact lens wearers, and those having had laser refractive surgery, intraocular refractive surgery, or eye disease and investigated with Rasch analysis and traditional statistics. Validity and reliability were assessed by Rasch fit statistics, principal components analysis (PCA), person separation, differential item functioning (DIF), item targeting, construct validity (correlation with visual acuity, contrast sensitivity, total root mean square [RMS] higher order aberrations [HOA]), and test–retest reliability (two-way random intraclass correlation coefficients [ICC] and 95% repeatability coefficients [R_c]).

Results. Rasch analysis demonstrated good precision, reliability, and internal consistency for all three scales (mean square infit and outfit within 0.81–1.27; PCA >60% variance explained by the principal component; person separation 2.08, 2.10, and 2.01 respectively; and minimal DIF). Construct validity was indicated by strong correlations with visual acuity, contrast sensitivity and RMS HOA. Test–retest reliability was evidenced by a minimum ICC of 0.867 and a minimum 95% R_c of 1.55 units.

Conclusions. The QoV Questionnaire consists of a Rasch-tested, linear-scaled, 30-item instrument on three scales providing a QoV score in terms of symptom frequency, severity, and bothersome. It is suitable for measuring QoV in patients with all types of refractive correction, eye surgery, and eye disease that cause QoV problems. (*Invest Ophthalmol Vis Sci.* 2010;51:5537–5545) DOI:10.1167/iovs.10-5341

Quality of vision (QoV) is a subjective entity based on an individual's unique perception of his or her vision. This perception is multifactorial, consisting not only of visual factors but also of psychological factors. Although optics and vision can be easily measured, none of these measurements explain how the patient subjectively perceives his or her vision; they provide only an indication.[1] Two patients may have identical visual function in terms of objective and subjective testing but very different perception of their QoV. Eye disease, refractive surgery, spectacles, and contact lenses may change QoV.[2–9] Therefore, the patient's perception of QoV may be an important outcome measure, but measuring perception requires a thoroughly developed and validated questionnaire.

Many vision-related questionnaires have been developed, but none that specifically measures only QoV. Some instruments include QoV questions, but the questions are mixed in with the measurement of other latent traits such as visual disability.[10–14] This combining of traits is a problem because all items that are combined to produce a score should measure a single trait; otherwise, the meaning of the measurement is unclear.[15] Second, averaging scores across all items erroneously assumes that the result provides an interval scale. It cannot be assumed that the difficulty of all questions is the same and the difficulty step between each category is constant; hence, the scaling may not be additive or linearly related to the trait under investigation.[16] Such instruments are inadequate for dealing with unanswered items and are not suitable to arithmetic functions. Item response models such as Rasch analysis have demonstrated the limitations in traditional summary (Likert) scoring[17–20] and overcome these problems with the transformation of ordinal raw scores into linear interval scales.[21]

Rasch analysis consists of a family of psychometric models that provides a valid measurement of the latent trait, in this case, QoV, with recognized superiority over summary scoring methods.[17,22] Rasch analysis provides a transformation to interval scoring via the probabilistic relationships between items (questions) and respondents. It has been used extensively in questionnaire development[23–26] and also for the redevelopment and improvement of preexisting questionnaires.[27–36]

Owing to the absence of a valid QoV instrument, a new instrument was conceived: the QoV Questionnaire, developed using Rasch analysis. The purpose of this study is to develop an instrument to measure QoV and to test its validity and reliability. The intended population includes patients with and without refractive correction in the form of spectacles, contact lenses, laser refractive surgery, and intraocular refractive surgery with various types of intraocular lenses (IOLs) and patients who have eye diseases that cause QoV problems, such as cataract.

Methods

Prestudy and Questionnaire Design

Items were identified via an extensive literature review and in focus groups with nonexperts and experts in the fields of refractive correction, questionnaire design, and subject interviews, to ensure content validity. The items were constructed and written at a comprehension level suitable for a 12-year-old.[37] Further focus groups and interviews helped assess the instrument for item redundancy, representation, and

From the [1]School of Biomedical Sciences, University of Ulster, Coleraine, United Kingdom; the [2]Cathedral Eye Clinic, University of Ulster, Belfast, Northern Ireland, United Kingdom; the [3]NH&MRC (National Health and Medical Research Council) Centre for Clinical Eye Research, Department of Optometry and Vision Science, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, South Australia, Australia; and the [4]Mater Hospital, Belfast Health and Social Care Trust, Belfast, Northern Ireland, United Kingdom.

Supported by the Department for Employment and Learning (DEL), Belfast, Northern Ireland, United Kingdom.

Submitted for publication February 8, 2010; revised April 12 and May 5, 2010; accepted May 5, 2010.

Disclosure: **C. McAlinden**, None; **K. Pesudovs**, None; **J.E. Moore**, None

Corresponding author: Colm McAlinden, BSc(Hons), MSc, PhD, School of Biomedical Sciences, University of Ulster, Cromore Road, Coleraine BT52 1SA, UK; colm.mcalinden@gmail.com.

TABLE 1. Questionnaire Consisting of 10 Items, Each with Three Questions Regarding the Frequency, Severity, and Bothersome, Resulting in a 30-Item Instrument

Item Number	Question
1	How often do you experience glare?
2	How severe is the glare?
3	How bothersome is the glare?
4	How often do you experience haloes?
5	How severe are the haloes?
6	How bothersome are the haloes?
7	How often do you experience starbursts?
8	How severe are the starbursts?
9	How bothersome are the starbursts?
10	How often do you experience hazy vision?
11	How severe is the hazy vision?
12	How bothersome is the hazy vision?
13	How often do you experience blurred vision?
14	How severe is the blurred vision?
15	How bothersome is the blurred vision?
16	How often do you experience distortion?
17	How severe is the distortion?
18	How bothersome is the distortion?
19	How often do you experience double or multiple images?
20	How severe are the double or multiple images?
21	How bothersome are the double or multiple images?
22	How often do you experience a fluctuation in your vision?
23	How severe is the fluctuation in your vision?
24	How bothersome is the fluctuation in your vision?
25	How often do you experience focusing difficulties?
26	How severe are the focusing difficulties?
27	How bothersome are the focusing difficulties?
28	How often do you experience difficulty judging distance or depth perception?
29	How severe is the difficulty judging distance or depth perception?
30	How bothersome is the difficulty judging distance or depth perception?

face validity, thus condensing the pilot questionnaire into a 10-symptom instrument with each item consisting of three questions regarding the frequency, severity, and bothersome of the items, hence, resulting in a 30-item QoV instrument (Table 1). These steps are displayed in a flowchart (Fig. 1). For the first seven QoV symptoms, an accompanying image was developed to aid in understanding the questions and to reduce the possibility of inconsistent responses (Fig. 2). Further focus groups and interviews helped optimize instrument layout, wording, and instructions.

Four response categories were chosen for each item, to provide the opportunity to discriminate levels of symptoms with a low possibility of redundancy, as it has been shown that respondents tend to use only four or five categories.[38] Categories were also labeled with descriptive wording that was refined with focus groups and interviews, with consistent wording for each item asking about frequency, severity, and bothersome (Table 2). Most of the focus group participants and interviewees preferred descriptive words rather than a numeric 0 to 3 response scale.

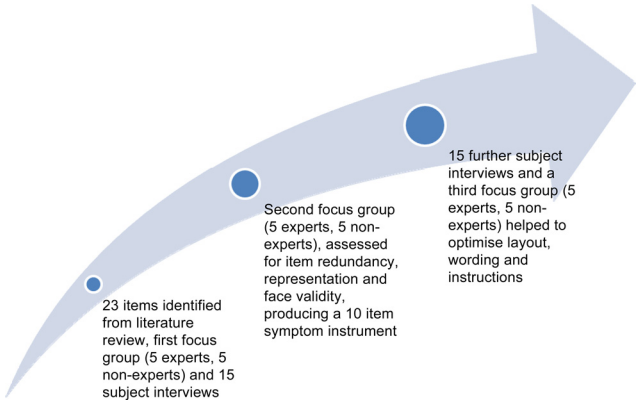

FIGURE 1. Flowchart demonstrating the steps involved in the questionnaire development.

Subjects

Subjects consisted of patients with and without refractive correction in the form of spectacles, contact lenses, laser refractive surgery, and intraocular refractive surgery with various types of IOLs (including multifocal lenses) and patients with cataract. The pilot questionnaire was completed by 900 respondents (mean age: 34 years; range, 21–78 years; percentage female: 57%). From this group, 150 respondents were spectacle wearers, 150 were contact lens wearers, 300 had undergone laser refractive surgery (consisting of laser in situ keratomileusis [LASIK], laser-assisted subepithelial keratectomy [LASEK], and photorefractive keratectomy [PRK] surgeries for various refractive errors), 150 had cataract, and 150 had undergone lens implantation surgery (with monofocal, multifocal, and pseudoaccommodative IOLs). These study groups were chosen because they were thought to represent the groups for which a QoV instrument would be most applicable. Equal numbers of subjects were selected for each group to ensure equal input into the formation of the questionnaire; so as not to bias the instrument in favor of one group or another. Statistical analysis was calculated with commercial software (Excel; Microsoft Corp., Redmond, WA and Statistical Package for the Social Sciences (SPSS), Version 17.0, Chicago, IL). The study was approved by the University of Ulster Research Ethics Committee and adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from the subjects after explanation of the nature and possible consequences of the study.

Rasch Analysis

The Rasch model is based on a probabilistic relationship between item difficulty and person ability. In a simple two-answer (endorse/

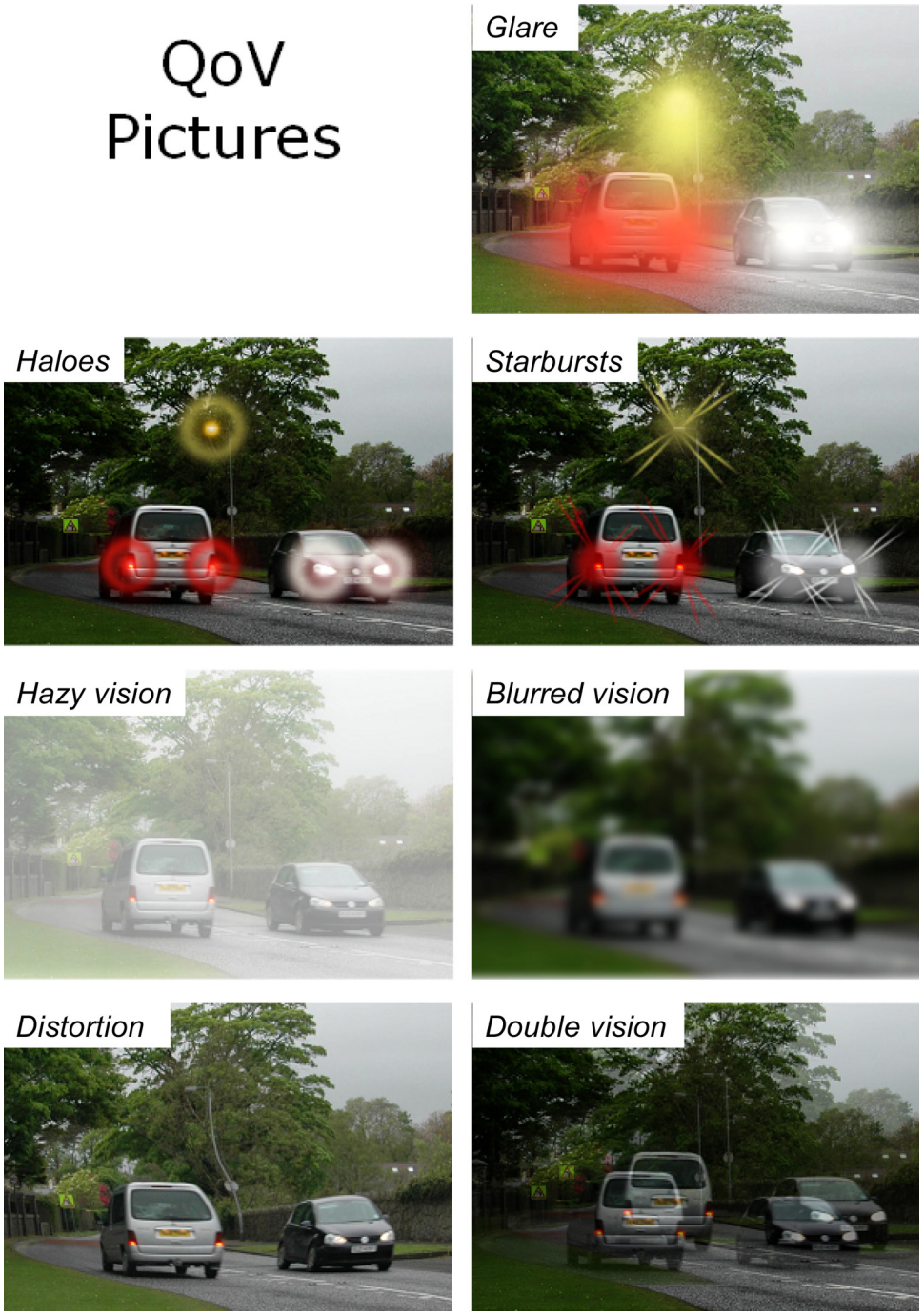

FIGURE 2. Quality of vision pictures used in the first seven visual symptom items.

does not endorse) model, the probability of endorsing an item is expressed as a function of the size of the difference between the ability (B) of the person (n) and the difficulty (D) of the item (i). Raw scores are converted into odds of success; the ratio of person percentage success (p) to person percentage failure ($1 - p$). The natural log of this ratio is the person ability estimate (Bn) and similarly for item difficulty estimates (Di). The result gives both person ability (Bn) and item difficulty (Di), expressed on a logit scale with the average logit of 0. Positive logits hence indicate higher than average probabilities of endorsing items and negative

TABLE 2. Four Response Categories Labeled with Descriptive Wording

Question Type	Response Category			
Frequency	Never (0)	Occasionally (1)	Quite often (2)	Very often (3)
Severity	Not at all (0)	Mild (1)	Moderate (2)	Severe (3)
Bothersome	Not at all (0)	A little (1)	Quite (2)	Very (3)

logits lower than average probability of endorsing items. This transformation turns the original questionnaire responses (raw ordinal data) into continuous interval data (like a ruler), providing a linear measurement. In the context of this study, difficulty and ability are synonymous with symptoms. A polytomous Andrich rating scale model[39] was performed using Bond and Fox Steps software (a customized version of Winsteps, Chicago, IL).[40] Three separate Rasch analyses were performed on the 30 items and 900 respondents: one analysis for the 10 frequency items, the 10 severity items, and the 10 bothersome items.

The response category performance was evaluated by observing whether the category calibration increased in an orderly fashion in the probability curves. Fit statistics (infit and outfit) were used to determine whether items fitted the Rasch model and whether the items measured a single latent trait (unidimensionality). Fit statistics focus on two aspects that can be reported as a mean square (MNSQ) or as a *z*-score (ZSTD), with expected values of 1 and 0, respectively. The MNSQ residual statistic is normalized to the average expected variance, such that a residual of less than 0.70 indicates at least 30% less variance than expected, suggesting possible redundancy or lack of variance to provide new information to the instrument, and residuals greater than 1.30 indicate at least 30% more variance than expected, suggesting that items may be measuring something different to the overall scale.[15,26,40] The MNSQ was used in this study, and an acceptable infit and outfit guidance for item removal is outside the range of 0.70 to 1.30, which was adhered to in this study.[15]

To complement fit statistics in the assessment of unidimensionality, we performed a principal components analysis (PCA) of the residuals.[41] Empiric and modeled variance explained by the measures were compared for the three scales. A variance greater than 60% indicates a low possibility of finding additional components. Multidimensionality was found to occur if a contrast had the strength of at least two items (eigenvalue >2.0), as above this value is the magnitude that occurs with random data.[40,42]

Rasch-derived person and item separation statistics indicate the overall precision of the instrument. The greater the value of person separation, the greater the precision, enabling a greater distinction between levels of symptoms.[43] A minimum acceptable cutoff for the person and item separation ratio was 2.0.[15]

Targeting was assessed to compare the item QoV score to person QoV score by observation of the item–person map. The logit scale was transformed into a 0 to 100 scale to produce the item–person map. The calibration of the items should be comparable across the five different groups included in this study; items operating in a similar way regardless of the group being investigated. The differences in item difficulty across respondent groups are known as differential item functioning (DIF).[44] The five groups assessed for DIF were spectacle wearers (1), contact lenses wearers (2), laser refractive surgery (3), cataracts (4), and intraocular lens or cataract surgery (5). A notable DIF was classified as a value greater than 1.0 logits.[31]

Instrument Performance Statistics

Construct validity refers to the ability of the instrument to measure QoV, which was assessed by correlations. Three construct hypotheses were evaluated due to the lack of a gold standard QoV instrument. The QoV scores in 20 randomly selected subjects was compared with logMAR visual acuity, contrast sensitivity, and total root mean square (RMS) higher order aberrations (HOA) by a Pearson product-moment correlation coefficient (*r*). LogMAR visual acuity and Pelli-Robson contrast sensitivity were measured on the Test Chart 2000 (Thomson Software Solutions, Herts, UK) under photopic conditions. The RMS HOA were measured with an aberrometer (OPD-Scan II ARK-10000; Nidek, Gamagori, Japan) across a 5-mm pupil diameter. Measurements were taken for both eyes, and the average value calculated and used for correlation analysis.

Twenty subjects were invited to repeat the questionnaire 10 days after the first completion, to assess the instrument's test–retest reliability, calculated by a two-way, single-measure intraclass correlation coefficient (ICC), and the repeatability was assessed by calculating the 95% repeatability coefficient (R_c).[45] The repeatability coefficient was calculated by determining the SD of the differences between repeated measures and multiplying by 2, which conforms to the British Standards Institution.[46] Statistical significance was set at $P < 0.05$.

A spreadsheet, entitled the QOV calculator, for converting raw scores to Rasch scaled scores (Excel; Microsoft) was developed for use by other investigators wishing to use the instrument and gain the benefits of Rasch scoring.

Results

Pilot Questionnaire Evaluation

The response categories functioned as intended, as illustrated by category structure calibration and observed averages increasing in an orderly fashion for the three scales, displayed graphically in Figure 3.

Fit statistics were all within the acceptable MNSQ range of 0.70 to 1.30 (Tables 3–5). For frequency items, the mean infit and outfit MNSQ statistic, respectively, was 1.00 ± 0.11 (range, 0.85–1.12) and 0.99 ± 0.12 (range, 0.83–1.20); for severity, 0.99 ± 0.09 (range, 0.84–1.13) and 1.00 ± 0.12 (range, 0.82–1.20); and for bothersome, 1.00 ± 0.11 (range, 0.85–1.27) and 0.99 ± 0.13 (range, 0.81–1.22).

Unidimensionality was also demonstrated with PCA. Empiric and modeled variance explained by the measures was similar for all three scales. All percentages were greater than 60%, indicating a low possibility of finding additional components. Unexplained variance explained by the first contrast was not greater than 2.0 eigenvalues for all three scales (Table 6).

Stable item symptoms were demonstrated with an item separation of 14.23 (reliability 1.00), 14.55 (reliability 1.00),

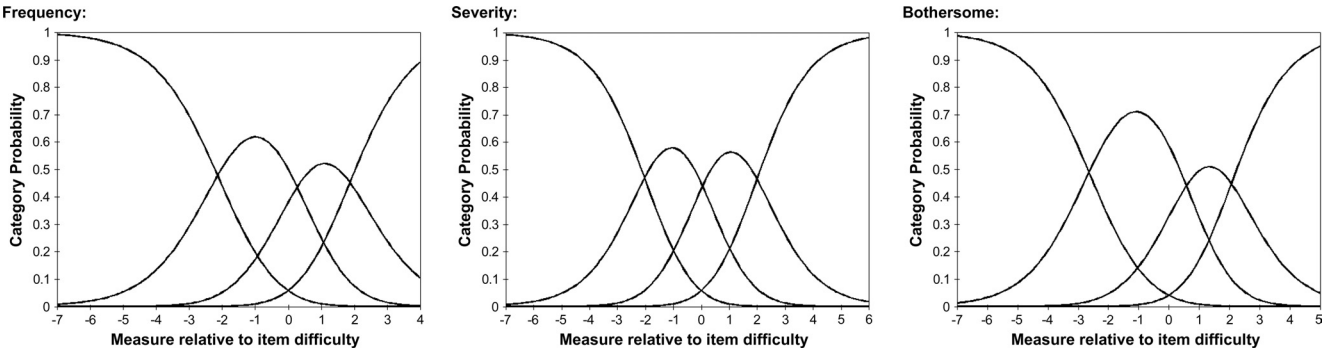

Figure 3. Category probability curves for the four response categories for frequency, severity and bothersome. The *x*-axis represents the severity of symptoms (difference between item and person calibration) and the *y*-axis represents the probability of the category being chosen.

TABLE 3. Results of the Frequency Scale Rasch Analysis

Item Number	Item Description	Infit MNSQ	Outfit MNSQ	SE	Symptoms (0–100 Scale)
1	Glare	1.01	1.02	0.43	37.32
4	Haloes	1.11	1.11	0.47	45.51
7	Starbursts	1.10	1.08	0.45	42.48
10	Hazy vision	1.11	1.07	0.48	47.47
13	Blurred vision	0.85	0.87	0.44	40.77
16	Distortion	0.90	0.88	0.53	53.32
19	Double vision	0.91	0.83	0.64	60.78
22	Fluctuation	0.86	0.83	0.48	47.18
25	Focusing difficulties	1.03	1.02	0.49	49.09
27	Depth perception	1.12	1.20	0.61	59.14
	Mean	1.00	0.99	0.50	48.31
	SD	0.11	0.12	0.07	7.21

and 12.82 (reliability 0.99) for the frequency, severity, and bothersome scales, respectively. Acceptable person discriminate symptoms were demonstrated with a person separation of 2.08, 2.10, and 2.01 for frequency, severity, and bothersome scales, respectively.

There was very little DIF (Table 7), with only 8 of 30 items affected and all 8 items included only three symptoms: hazy vision, distortion, and double vision. In all cases, it was the spectacle- or contact lens–wearing group that had an item calibration different from that of one of the disease groups, usually the cataract or cataract surgery groups.

Figure 4 demonstrates the person–item map with subjects appearing on the left and items on the right in ascending order of more frequent, severe, or bothersome symptoms. The items are located in order of their impact on QoV from the bottom to the top. Greater scores indicate greater amounts of symptom frequency, severity, or bothersome. The map is the logit scale transformed into a 0 to 100 scale, with item targeting indicating items in all three scales had more symptoms than person symptoms. The mean mistargeting was 2.0 logits for frequency, 1.9 logits for severity, and 2.6 for bothersome.

Instrument Performance

Construct validity indicated positive correlation between QoV scores and logMAR visual acuity; frequency scale $r = 0.72$ (95% CI, 0.41–0.88, $P < 0.001$), severity scale $r = 0.64$ (95% CI, 0.28–0.85, $P = 0.002$), and bothersome scale $r = 0.35$ (95% CI, −0.09 to 0.70, $P = 0.130$). There was a negative correlation between QoV scores and contrast sensitivity; frequency scale $r = -0.80$ (95% CI, −0.56 to −0.92, $P < 0.001$), severity scale $r = -0.73$ (95% CI, −0.43 to −0.90, $P < 0.001$), and bothersome scale $r = -0.57$ (95% CI, −0.17 to −0.80, $P = 0.009$). A positive correlation was found between QoV scores and total HOA; frequency scale $r = 0.71$ (95% CI, 0.39–0.88, $P < 0.001$), severity scale $r = 0.61$ (95% CI, 0.22–0.83, $P = 0.005$), and bothersome scale $r = 0.69$ (95% CI, 0.36–0.87, $P = 0.001$).

Test–Retest Reliability and Repeatability

The two-way, single-measure ICC for test–retest reliability was 0.992 (95% CI, 0.981–0.997) for the frequency scale, 0.872 (95% CI, 0.851–0.901) for the severity scale, and 0.867 (95% CI, 0.833–0.898) for the bothersome scale. The 95% $R_c = 2.96$, 1.55, and 3.32 for the frequency, severity, and bothersome scales, respectively.

DISCUSSION

The QoV Questionnaire was developed using conventional statistics, and Rasch analysis providing a reliable and valid quantitative linear measurement. The major advantage of the use of Rasch analysis is that estimates are on a linear interval scale, not an ordinal scale, and so the QoV Questionnaire is capable of measuring the change in symptoms more accurately and is better equipped at dealing with omitted items, hence overcoming the limitations of other instruments.[10–14] The prestudy consisted of item selection from a wide variety of contributors across a variety of disciplines, including focus groups, subject interviews, and literature reviews ensuring that all appropriate items were included. The selection process was carefully refined with multiple pilot tests and further with focus groups and interviews. This process is important, as it provides content validity. The questionnaire was administered to a target population where it is anticipated it will be used.

The 30 item QoV Questionnaire was separated into three scales because there are three different types of questions asked: frequency of visual symptoms, the severity of the symp-

TABLE 4. Results of the Severity Scale Rasch Analysis

Item Number	Item Description	Infit MNSQ	Outfit MNSQ	SE	Symptoms (0–100 Scale)
2	Glare	1.06	1.10	0.42	37.16
5	Haloes	1.07	1.08	0.45	44.95
8	Starbursts	1.03	1.01	0.44	42.56
11	Hazy vision	1.13	1.08	0.47	48.12
14	Blurred vision	0.91	0.96	0.43	41.12
17	Distortion	0.84	0.85	0.52	53.79
20	Double vision	0.92	0.82	0.63	60.51
23	Fluctuation	0.88	0.84	0.46	46.91
26	Focusing difficulties	1.01	0.99	0.47	48.75
29	Depth perception	1.09	1.20	0.60	59.21
	Mean	0.99	1.00	0.49	48.31
	SD	0.09	0.12	0.07	7.21

TABLE 5. Results of the Bothersome Scale Rasch Analysis

Item Number	Item Description	Infit MNSQ	Outfit MNSQ	SE	Symptoms (0–100 Scale)
3	Glare	0.96	0.96	0.49	38.26
6	Haloes	0.97	0.97	0.53	45.72
9	Starbursts	1.03	1.02	0.51	43.24
12	Hazy vision	1.05	1.04	0.54	47.41
15	Blurred vision	0.85	0.85	0.50	40.14
18	Distortion	0.95	0.89	0.59	53.76
21	Double vision	0.88	0.81	0.69	60.57
24	Fluctuation	1.01	0.97	0.53	45.96
27	Focusing difficulties	1.27	1.22	0.54	47.97
30	Depth perception	1.06	1.22	0.68	60.03
	Mean	1.00	0.99	0.56	48.31
	SD	0.11	0.13	0.07	7.24

tom, and how bothered the patient is by the symptom. Ten frequency items, 10 severity items, and 10 bothersome items were formed, with the three scales undergoing three separate Rasch analyses. This provides for three measures of QoV: a measure of the frequency of symptoms, the severity of symptoms, and the level of bother of symptoms.

The results of the Rasch analysis demonstrate that the QoV Questionnaire has good psychometric qualities. Category probability curves indicated good category discrimination and orderly category structure for all three scales. Fit statistics indicated that all items fitted the Rasch model and together with PCA demonstrated unidimensionality. Good item and person separation values were also found for all three scales, indicating stable item difficulty and good person discriminative ability. There was minimal DIF, with the questions relating to hazy vision, distortion, and double vision, showing some differences in diseased and nondiseased eyes. However, the magnitude of the DIF was unlikely to be problematic. The person-item map indicates that items targeted the more symptomatic end of the QoV scales. Sometimes items are removed from questionnaires so that the average item difficulty/symptoms would correspond to the average person ability/symptoms. However, for a symptom instrument, it was anticipated that some mistargeting would occur, as many patients do not have any QoV problems, a situation frequently encountered in symptom-based questionnaires. Therefore, we feel the targeting is satisfactory.

In the assessment of the instrument's performance, the QoV scores were compared to three constructs; visual acuity, contrast sensitivity and total RMS HOA. Correlations were as expected, with positive correlations between QoV scores and logMAR visual acuity and total RMS HOA. A negative correlation was found for QoV scores with contrast sensitivity (better contrast sensitivity is a higher score, opposite to logMAR visual acuity and total RMS HOA). This questionnaire may measure QoV but differentiation between right and left eyes is not possible in these construct hypotheses; hence an average value was calculated for both eyes. We are aware that this is a simplification assuming equal integration of image perception of both eyes. A very high correlation would suggest that the instrument provides information which is too similar to the construct under investigation and not providing sufficient additional information. A very low correlation may indicate that the two measures which are hypothesized to be related are actually not well related. Suggested guidelines for correlation coefficients are in the range of 0.3 to 0.9, which was found in this study.[15] The QoV Questionnaire test–retest performance was excellent for all three scales with a minimum ICC of 0.867 and a minimum 95% R_c test–retest repeatability of 1.55 units.

Refractive correction is an extremely common treatment because of the large prevalence of refractive error in the population.[47] Refractive error is typically corrected with spectacles, contact lenses, laser refractive surgery, or intraocular refractive surgery. However, QoV symptoms are one of the most frequent problems after refractive surgery correction in its various forms and nonsurgical correction, such as spectacles and contact lenses.[3–5,48,49] In particular, patients undergoing laser refractive surgery tend to have excellent vision with spectacles and a decrease in QoV caused by laser surgery is ultimately undesirable. However, the incidence of QoV problems are fewer with modern ablation lasers that employ larger treatment/blend zones, eye trackers, and wavefront technology[2]; however, the problems are still very significant with intraocular refractive surgery, particularly with multifocal IOLs.[5,50] Spectacle lenses and contact lenses may affect QoV, depending on many factors such as the lens power, material, base curve, asphericity, refractive index, antireflective coatings, and lens scratches, with more recent developments using wavefront optics in an attempt to improve visual performance and QoV (Jethmalani J, et al. *IOVS* 2004;45:ARVO E-Abstract 2764).[6–8,51–53] Crystalline lens opacities have long been known to cause QoV problems.[9]

The cause of these symptoms may be ascertained from objective testing, but it is unknown how different visual aspects combine to provide overall visual quality perceptions.[54] It is also unknown how objective clinical test results relate to a patient's everyday life along with the variation among patients.[55] These issues are the focus of our ongoing research in this area. In light of this need, it is necessary to accurately evaluate the patient's subjective QoV, which is best achieved with a ques-

TABLE 6. PCA for the Three Scales

	Frequency		Severity		Bothersome	
	Empirically	Modeled	Empirically	Modeled	Empirically	Modeled
Variance explained by the measurement	67.1%	66.8%	68.2%	68.2%	63.2%	62.9%
Unexplained variance by the first contrast	1.9 Eigenvalue		2.0 Eigenvalue		1.9 Eigenvalue	

TABLE 7. Items Showing DIF across All Three Scales

Items	Groups				
	1 Spectacles Wearers	2 Contact Lens Wearers	3 Laser Refractive Surgery	4 Cataract	5 Intraocular Lens Or Cataract Surgery
10 (Hazy vision: frequency)	4 (1.01)	—	—	—	—
16 (Distortion: frequency)	5 (1.04)	—	—	—	—
19 (Double vision: frequency)	—	4 (1.02) 5 (1.07)	—	—	—
17 (Distortion: severity)	4 (1.02)	—	—	—	—
20 (Double vision: severity)	—	4 (1.19) 5 (1.24)	—	—	—
12 (Hazy vision: bothersome)	4 (1.03) 5 (1.06)	—	—	—	—
18 (Distortion: bothersome)	3 (1.03) 4 (1.08) 5 (1.34)	—	—	—	—
21 (Double vision: bothersome)	—	3 (1.21) 4 (1.4) 5 (1.25)	—	—	—

Data are expressed in logits (log-odds of the level of symptoms of an item relative to the symptoms of the total set of items analyzed) and the listed subgroup rated these items as easier relative to other symptoms by the amount of logits indicated in the parentheses.

tionnaire. The QoV Questionnaire provides a standardized measure of a patient's QoV perception before and after refractive correction or surgery, disease progression, medical therapy, or surgical intervention. It may be used before surgery to help decide the best refractive surgery procedure and after surgery to assess the effectiveness of the given procedure and to monitor changes in QoV. It would also provide insight into the effects of complications on QoV and different types of spectacle and contact lenses. The QoV Questionnaire provides an additional assessment in the efficacy of various refractive procedures such as the comparison between the different laser ablation algorithms when keratorefractive surgery is performed or the comparison of multifocal IOLs when intraocular refractive surgery is performed. Most funding bodies also insist on a patient-reported outcome assessment for a clinical trial of a disease treatment or intervention; hence, an instrument of this nature is warranted. An alternative approach to developing a questionnaire could have been to use an item bank and computer adaptive testing approach, as has been mooted for visual disability and quality of life.[56] This approach would be valid also and advantageous if there were a large number of QoV items (as there are for visual disability) because implementing a subset of items targeted to the person would be more efficient. However, since there appears to be only a small number of QoV items (10 concepts), we elected to take the simple questionnaire approach as we thought it may aid in utilization.

In conclusion, we have developed an instrument to measure subjective QoV: the QoV Questionnaire. It is a Rasch-tested, lin-

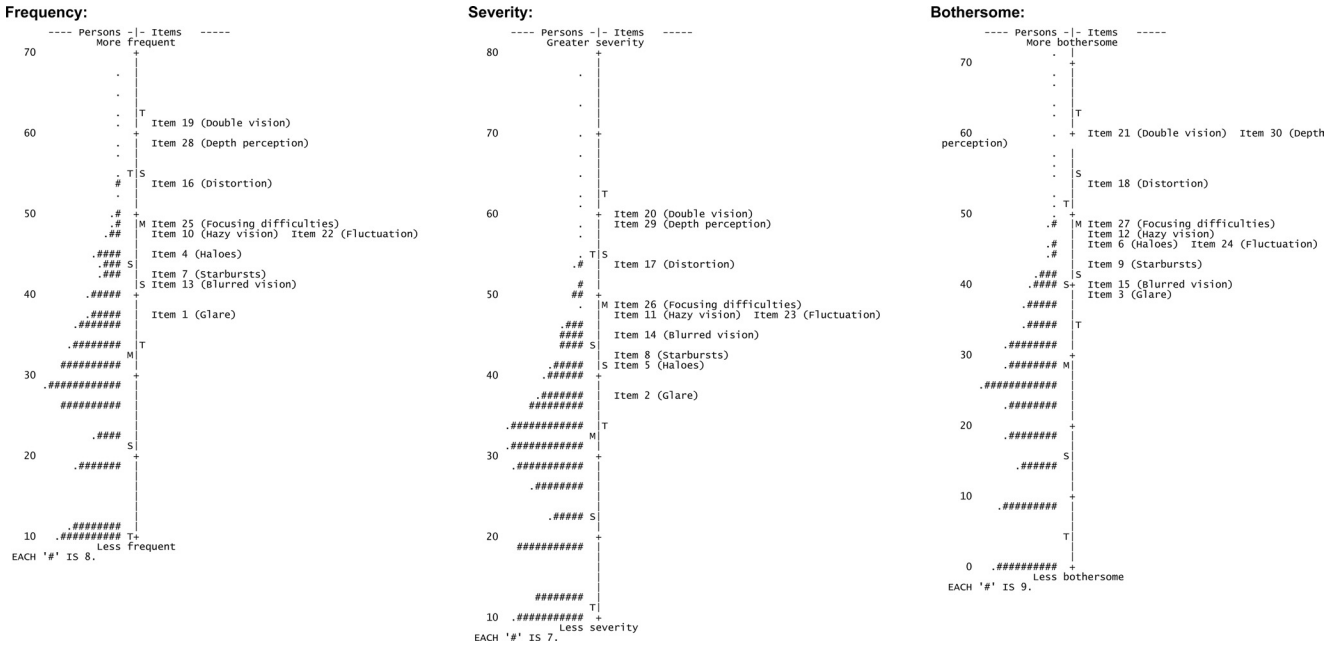

FIGURE 4. Person-item map for the 30-item QoV Questionnaire. Subjects appear on the left (each # in the person column represents three persons and each dot represents one to two persons).

ear-scaled, 30-item instrument on three separate scales providing a QoV score for frequency, severity and bothersome of symptoms. It is also equipped to quantify change and account for missing data. It is suitable for use in clinical practice, clinical trials, and research studies for measuring QoV in patients with and without refractive correction in the form of spectacles, contact lenses, laser refractive surgery, intraocular refractive surgery with various types of IOLs and for patients with eye disease, such as cataract.

A PDF file containing the QoV Questionnaire and a convenient Excel QoV calculator is available from the corresponding author.

Acknowledgments

The authors thank Trevor Bond, Michael Johnson, and Mike Linacre for helpful discussions on various aspects of questionnaire design and Rasch analysis.

References

1. McGhee CN, Craig JP, Sachdev N, Weed KH, Brown AD. Functional, psychological, and satisfaction outcomes of laser in situ keratomileusis for high myopia. *J Cataract Refract Surg.* 2000;26: 497–509.
2. Sakimoto T, Rosenblatt MI, Azar DT. Laser eye surgery for refractive errors. *Lancet.* 2006;367:1432–1447.
3. O'Brart DP, Lohmann CP, Fitzke FW, Smith SE, Kerr-Muir MG, Marshall J. Night vision after excimer laser photorefractive keratectomy: haze and halos. *Eur J Ophthalmol.* 1994;4:43–51.
4. O'Brart DP, Lohmann CP, Fitzke FW, et al. Disturbances in night vision after excimer laser photorefractive keratectomy. *Eye* (*Lond*). 1994;8:46–51.
5. Woodward MA, Randleman JB, Stulting RD. Dissatisfaction after multifocal intraocular lens implantation. *J Cataract Refract Surg.* 2009;35:992–997.
6. Lohmann CP, Fitzke F, O'Brart D, Muir MK, Timberlake G, Marshall J. Corneal light scattering and visual performance in myopic individuals with spectacles, contact lenses, or excimer laser photorefractive keratectomy. *Am J Ophthalmol.* 1993;115:444–453.
7. Lohmann CP, Fitzke FW, O'Brart D, Muir MK, Marshall J. Halos–a problem for all myopes?—a comparison between spectacles, contact lenses, and photorefractive keratectomy. *Refract Corneal Surg.* 1993;9:S72–S75.
8. Rubin ML. Spectacles: past, present, and future. *Surv Ophthalmol.* 1986;30:321–327.
9. Alio JL, Schimchak P, Negri HP, Montes-Mico R. Crystalline lens optical dysfunction through aging. *Ophthalmology.* 2005;112: 2022–2029.
10. Fraenkel G, Comaish F, Lawless MA, et al. Development of a questionnaire to assess subjective vision score in myopes seeking refractive surgery. *J Refract Surg.* 2004;20:10–19.
11. Brunette I, Gresset J, Boivin JF, Boisjoly H, Makni H. Functional outcome and satisfaction after photorefractive keratectomy. part 1: Development and validation of a survey questionnaire. *Ophthalmology.* 2000;107:1783–1789.
12. Brunette I, Gresset J, Boivin JF, et al. Functional outcome and satisfaction after photorefractive keratectomy, part 2: survey of 690 patients. *Ophthalmology.* 2000;107:1790–1796.
13. Schein OD. The measurement of patient-reported outcomes of refractive surgery: the refractive status and vision profile. *Trans Am Ophthalmol Soc.* 2000;98:439–469.
14. Hays RD, Mangione CM, Ellwein L, Lindblad AS, Spritzer KL, McDonnell PJ. Psychometric properties of the national eye institute-refractive error quality of life instrument. *Ophthalmology.* 2003;110:2292–2301.
15. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci.* 2007; 84:663–674.
16. Townsend JT, Ashby FG. Measurement scales and statistics: the misconception misconceived. *Psychol Bull.* 1984;96:394–401.
17. Massof RW. The measurement of vision disability. *OptomVisSci.* 2002;79:516–552.
18. Pesudovs K. Patient-centered measurement in ophthalmology: a paradigm shift. *BMC Ophthalmol.* 2006;6:25.
19. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989; 70:857–860.
20. Fisher WP Jr, Eubanks RL, Marier RL. Equating the MOS SF36 and the LSU HSI physical functioning scales. *J Outcome Meas.* 1997; 1:329–362.
21. 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–I36.
22. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research; 1960.
23. Pesudovs K, Garamendi E, Elliott DB. The quality of life impact of refractive correction (QIRC) questionnaire: development and validation. *Optom Vis Sci.* 2004;81:769–777.
24. Pesudovs K, Garamendi E, Elliott DB. The contact lens impact on quality of life (CLIQ) questionnaire: development and validation. *Invest Ophthalmol Vis Sci.* 2006;47:2789–2796.
25. Gupta N, Wolffsohn JS, Naroo SA, Davies LN, Gibson GA, Shah S. Development of a near activity visual questionnaire to assess accommodating intraocular lenses. *Cont Lens Anterior Eye.* 2007; 30:134–143.
26. Johnson ME, Murphy PJ. Measurement of ocular surface irritation on a linear interval scale with the ocular comfort index. *Invest Ophthalmol Vis Sci.* 2007;48:4451–4458.
27. 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.
28. Pesudovs K, Caudle LE, Rees G, Lamoureux EL. Validity of a visual impairment questionnaire in measuring cataract surgery outcomes. *J Cataract Refract Surg.* 2008;34:925–933.
29. Lamoureux EL, Pesudovs K, Pallant JF, et al. 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;15:224–233.
30. Lamoureux EL, Pesudovs K, Thumboo J, Saw SM, Wong TY. An evaluation of the reliability and validity of the visual functioning questionnaire (VF-11) using Rasch analysis in an Asian population. *Invest Ophthalmol Vis Sci.* 2009;50:2607–2613.
31. Lundstrom M, Pesudovs K. Catquest-9SF patient outcomes questionnaire: Nine-item short-form Rasch-scaled revision of the Catquest Questionnaire. *J Cataract Refract Surg.* 2009;35:504–513.
32. Pesudovs K, Gothwal VK, Wright TA. Visual disability assessment: valid measurement of activity limitation and mobility in cataract patients. *Br J Ophthalmol.* 2010;94:777–781.
33. Gothwal VK, Wright T, Lamoureux EL, Pesudovs K. Psychometric properties of visual functioning index using Rasch analysis. *Acta Ophthalmol.* Published online June 26, 2009.
34. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Using Rasch analysis to revisit the validity of the cataract TyPE spec instrument for measuring cataract surgery outcomes. *J Cataract Refract Surg.* 2009;35:1509–1517.
35. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. The impact of cataract surgery questionnaire: re-evaluation using Rasch analysis. *Acta Ophthalmol.* Published online October 30, 2009.
36. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Cataract symptom score questionnaire: Rasch revalidation. *Ophthalmic Epidemiol.* 2009;16:296–303.
37. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use.* Oxford, UK: Oxford University Press; 2003.
38. Thomee R, Grimby G, Wright BD, Linacre JM. Rasch analysis of visual analog scale measurements before and after treatment of patellofemoral pain syndrome in women. *Scand J Rehabil Med.* 1995;27:145–151.

39. Andrich D. A rating scale formulation for ordered response categories. *Psychometrika.* 1978;43:561–573.
40. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates; 2007.
41. Wright BD. Local dependency, correlations and principal components. *Rasch Meas Trans.* 1996;10:509–511.
42. Linacre JM. *A User's Guide to WINSTEPS*. Winsteps.Com 2009.
43. Mallinson T. Why measurement matters for measuring patient vision outcomes. *Optom Vis Sci.* 2007;84:675–682.
44. Holland PW, Wainer H. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993.
45. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med.* 1990;20:337–340.
46. British Standards Institution. *Precision of Test Methods. 1: Guide for the Determination and Reproducibility of a Standard Test Method* (BS5497, part 1). London: BSI; 1979.
47. Vitale S, Ellwein L, Cotch MF, Ferris FL 3rd, Sperduto R. Prevalence of refractive error in the united states, 1999–2004. *Arch Ophthalmol.* 2008;126:1111–1119.
48. Pop M, Payette Y. Risk factors for night vision complaints after LASIK for myopia. *Ophthalmology.* 2004;111:3–10.
49. Uusitalo RJ, Aine E, Sen NH, Laatikainen L. Implantable contact lens for high myopia. *J Cataract Refract Surg.* 2002;28:29–36.
50. Leyland M, Pringle E. Multifocal versus monofocal intraocular lenses after cataract extraction. *Cochrane Database Syst Rev.* 2006;(4):CD003169.
51. Margrain TH, Owen C. The misting characteristics of spectacle lenses. *Ophthalmic Physiol Opt.* 1996;16:108–114.
52. Jethmalani J, Chomyn J, Abdel-Sadek G, et al. Wavefront guided spectacle lenses for emmetropes and myopes. *Invest Ophthalmol Vis Sci.* 2004;45:2764.
53. Lopez-Gil N, Castejon-Mochon JF, Benito A, et al. Aberration generation by contact lenses with aspheric and asymmetric surfaces. *J Refract Surg.* 2002;18:S603–S609.
54. McAlinden C, Moore JE. Higher order aberrations using the NIDEK OPD-scan and AMO WaveScan. *J Refract Surg.* 2010;26:605–608.
55. McAlinden C, Moore JE. Comparison of higher order aberrations after LASIK and LASEK for myopia. *J Refract Surg.* 2010;26:45–51.
56. Pesudovs K. Item banking: a generational change in patient-reported outcome measurement. *Optom Vis Sci.* 2010;87(4):285–293.