1040-5488/04/8110-0769/0 VOL. 81, NO. 10, PP. 769–777
OPTOMETRY AND VISION SCIENCE

ORIGINAL ARTICLE

The Quality of Life Impact of Refractive Correction (QIRC) Questionnaire: Development and Validation

KONRAD PESUDOVS, PhD, FAAO, ESTIBALIZ GARAMENDI, MSc, and DAVID B. ELLIOTT, PhD, FAAO

Department of Optometry, University of Bradford, Bradford, West Yorkshire, United Kingdom

ABSTRACT: *Background.* The purpose of the study was to develop a questionnaire that could quantify the quality of life (QOL) of people with refractive correction by spectacles, contact lenses, and refractive surgery in the prepresbyopic age group. *Methods.* The questionnaire was developed and validated using traditional methods and Rasch analysis. A 90-item pilot questionnaire was developed through extensive literature search and use of professional and lay focus groups. Pilot study data were obtained from 306 subjects for item reduction to produce the 20-item Quality of Life Impact of Refractive Correction (QIRC) questionnaire. Validity and reliability studies (test-retest reliability with intraclass correlation coefficient and Bland-Altman limits of agreement, and internal consistency with Rasch fit statistics, factor analysis, and Cronbach's α) were performed from data of an additional 312 subjects. *Results.* Rasch analysis demonstrated QIRC has good precision, reliability, and internal consistency (person separation, 2.03; reliability, 0.80; root-mean-square measurement error, 3.25; mean square ± SD infit, 0.99 ± 0.38; outfit, 1.00 ± 0.39; item infit range, 0.70 to 1.24; and item outfit range, 0.78 to 1.32). The items (mean score, 50.3 ± 7.3) were well targeted to the subjects (mean score, 47.8 ± 5.5) with a mean difference of 2.45 (scale range, 0 to 100) units. Test-retest reliability (intraclass correlation coefficient, 0.88; coefficient of repeatability, ±6.85 units), factor loading range (0.40 to 0.76), and Cronbach's α (0.78) also indicated the reliability and validity of QIRC. *Conclusions.* The 20-item QIRC questionnaire, which quantifies the QOL of people with refractive correction by spectacles, contact lenses, and refractive surgery in the prepresbyopic age group, was developed using Rasch analysis and shown to be valid and reliable. The use of Rasch scaling allows scores to be treated as a valid continuous variable. QIRC has broad applicability for cross-sectional and outcomes research. (Optom Vis Sci 2004;81:769–777)

Key Words: quality of life, questionnaire, Rasch analysis, reliability and validity

The purpose of this study was to develop a questionnaire to measure the quality of life (QOL) impact of spectacles, contact lenses, and refractive surgery. The need to rate the status of patients using a measure of QOL is well established for clinical research and practice, and a large number of QOL questionnaires have been developed.[1] This is important for refractive error correction, which in the United States is a $22.8 billion industry, with 59% of the US population possessing a refractive correction.[2] Although spectacles dominate this, about 12% of the adult population wears contact lenses, and 6.1 million (2.2%) have had refractive surgery, including 1.2 million (~0.4%) in 2002.[2]

Two formal, conventionally validated questionnaires have been developed for refractive surgery outcomes: The National Eye Institute Refractive Error Quality of Life Instrument (NEI-RQL)[3] and the Refractive Status Vision Profile (rapid serial visual presentation [RSVP]).[4] The RSVP was developed almost exclusively on a refractive surgery population (92% of subjects); therefore, it is only valid for refractive surgery.[4] The RSVP and the NEI-VFQ have been shown to be insensitive to QOL issues relevant to people wearing contact lenses.[5, 6] The development and validation of the NEI-RQL were spread across four articles, and despite rigorous work with focus groups, there is no report of how the 42 items were selected for the final questionnaire.[3, 7–9] However, independent analysis does confirm that the NEI-RQL can discriminate between modes of refractive correction.[10] Other studies that report QOL issues before and after refractive surgery have used informal, nonvalidated questionnaires.[11–13] One little known article has shown differences in QOL between spectacle and contact lens wearers,[14] but this was restricted to psychosocial issues.

An important issue is the scoring strategy used for questionnaires. Many currently available questionnaires, such as the RSVP and NEI-RQL, use traditional Likert (summary) scoring,[15] and Rasch analysis has been used to highlight its superiority to Likert scoring in vision-related questionnaires.[16, 17] The problems are that these scales assume that equal distances between response choices represent equal distances in the dimension measured and that items represent equal difficulty and should therefore be scored equally. For example, in a visual disability questionnaire, such as the Activities of Daily Vision Scale,[18] does a response of "a little difficulty" (score of 4) represent a level of ability twice as good as "extreme difficulty" (score of 2), and is that twice as good again as "unable to perform the activity due to vision" (score of 1)? Logic suggests this is clearly not the case, and Rasch analysis has been used to confirm this.[17] Similarly, should an answer of "a little difficulty" to the question regarding visual difficulties "driving at night" score the same as the "a little difficulty" with "driving during the day?" Once again, common sense indicates that driving at night is a more difficult visual task, and this has been confirmed by Rasch analysis, and appropriate weightings have been determined.[17] Rasch analysis shows that all the visual ability tasks vary in difficulty and can be scored accordingly, and the same is true for items in a QOL scale.[19] The Likert scoring system, with one to five (or similar) responses and scoring all the questions the same regardless of the difficulty of the task, is wholly illogical and has been shown to be so by Rasch analysis.[17, 19–21] Rasch analysis provides a method to test scale assumptions and then modify the scale structure so that the output variable is scaled as a valid continuous measure. Rasch analysis is also helpful for questionnaire development because it provides powerful insight into internal consistency and the targeting of items to subjects (the extent to which item difficulty suits population ability). Therefore, we developed and validated a questionnaire, using Rasch analysis, for the measurement of the impact of refractive correction on the QOL: The Quality of Life Impact of Refractive Correction (QIRC) questionnaire. The questionnaire was targeted at adults requiring refractive correction who did not have other ophthalmic problems but confined to the prepresbyopic population because the majority of contact lens and refractive surgery patients are prepresbyopes, and presbyopes are likely to encounter different issues than prepresbyopes related to the use of reading glasses, bifocals, and varifocals and the much higher prevalence of ocular disease.

METHODS

The QIRC questionnaire was developed following standard procedures[22] and augmented by item reduction and scoring applications of Rasch analysis.[23, 24] Rasch analysis was performed using Winsteps (version 3.35, Chicago, IL),[25] which calculates Wright and Masters'[23] version of the Rasch model estimates using joint maximum likelihood estimation. The Rasch model does not assume values for response categories (e.g., 1, 2, 3, and so on) but does assume that all the categories lie on the same scale. The Rasch model gives the probability of selecting a particular response category in terms of the interaction between "response severity" and subject measure (in this case, refractive error-related QOL) through an iterative logistic process.[23] The resulting response scale calibrations and person measures are expressed in log-odd units (natural logarithm of an odds ratio), or logits, positioned along a hierarchical scale with logits of greater magnitude representing increasing vision-related QOL.[26] The validity of the assumption that all the items are on the same scale is tested and reported in terms of model fit statistics. This is important for item reduction (see below).

Domain and Item Identification and Selection

Domains and items thought likely to be influenced by refractive correction were drawn from six sources: a search of the general QOL literature;[27–31] a search of the vision-specific QOL literature,[16, 32, 33] including that relating to refractive error correction;[4, 12, 34–38] a search of the QOL in cosmetic surgery literature;[39, 40] retrospective analysis of case records at the University of Bradford Eye Clinic; invited responses from 63 practitioners and allied health workers in the fields of optometry, ophthalmology, contact lens practice, refractive surgery, and psychology; and focus groups (lay people and professionals in the aforementioned fields). To ensure content validity, 647 items (broken down by domain: physical functioning, 176; health concerns, 19; well-being, 192 [of these, psychological well-being, 108; social well-being, 74], convenience issues, 85; symptoms, 97; economic issues, 54; and cognitive issues, 24) were identified. These were reduced to 115 items by the professional focus groups. In this process, many items were merged for having similar content (e.g., reading small print and reading medicine bottles). Others were discarded for not being relevant to a majority of people (e.g., ability to cross-stitch). These 115 items were formatted into a self-administration questionnaire for additional discussion by lay focus groups. Question format was kept as regular as possible, but different content areas required different question syntax. Two styles of questions were chosen: severity assessment (e.g., How much difficulty do you have. . .?) and incidence (e.g., During the past month, how often have you experienced. . .?). A 5-category response scale was chosen because it has been shown to be more useful and easier to complete compared with 4- and 7-category response scales and a visual analogue scale.[41] Suitably spaced response labels were selected from the research literature.[42]

The three lay focus groups (one group each of spectacle wearers, contact lens wearers, and postrefractive surgery patients) were asked to assess each item for the likely impact of refractive correction, relevance to a majority of subjects, and ease of understanding. The lay focus groups recommended discarding an additional 25 items and numerous minor rewordings of instructions and questions to assist comprehension. Advice was also taken from the lay focus groups on the wording of the instructions. The 90 items for the pilot questionnaire were distributed among the domains of visual function (19), health concerns (8), well-being (16), convenience issues (20), symptoms (16), economic issues (8), and cognitive issues (3).

90-Item Pilot Questionnaire

The study was designed so that the final questionnaire would be relevant to the population of the United Kingdom. Thus, it was administered in 18 centers throughout the United Kingdom, with gender, ethnicity, and socioeconomic classification data collected.

The centers were chosen to provide data from rural and urban areas in the United Kingdom and with a good geographical spread.

Subjects were chosen on a consecutive patient basis under the constraints of time inherent in a commercial practice. The success of recruiting a representative population was established through comparison of the responses with these demographic questions against UK national data.[43] Inclusion criteria were age between 16 and 35 years (adult prepresbyopic age), myopia, hyperopia, or astigmatism managed with spectacles, contact lenses, or by refractive surgery at least 1 year previously. Exclusion criteria were previous ocular (other than refractive) surgery, eye disease, neurologic disease, systemic disease, or medication that might alter visual function and inability to read or understand the questionnaire. Informed consent was obtained from all the subjects after the nature of the study had been fully explained. The tenets of the Declaration of Helsinki were followed, and the study gained approval from the university ethical committee. The 90-item pilot questionnaire was completed by self-administration by 344 subjects across settings, including optometry, contact lens, and refractive surgery (mostly, but not exclusively laser-assisted *in situ* keratomileusis) practices.

Twelve questionnaires were discarded because of absent demographic data or >33% missing item responses. Rasch analysis was used to identify unusual response patterns. The Rasch model fit statistics, infit and outfit mean square, which compare the predicted responses with the observed, were used to monitor the compatibility of the data with the model. Outfit (outlier sensitive fit) mean square is the conventional sum of squared standardized residuals and is sensitive to occasional responses that are different to the expected response. For infit (information-weighted fit) mean square, each squared standardized residual value is first weighted by its variance and then summed. In this way, infit takes less notice of extreme responses because it is weighted to be sensitive to responses that are close to a respondent's level of (in this case) refractive error-related QOL. Infit and outfit mean squares have an expected value of 1, with those <0.80 representing items that over-fit the model and are too predictable (they have at least 20% less variation than was expected). Over-fitting items may be redundant or noncontributory because they lack variance. Mean squares >1.20 represent misfit (at least 20% more variance than was expected), and values >1.40 represent substantial misfitting. A high item infit or outfit suggests that the item measures something different to the overall scale.[44] Sixty-five subjects gave poor Rasch fit statistics (outfit and infit mean square, >1.40), indicating their responses were different from the majority of subjects; therefore, the authors reviewed their questionnaires. Forty-nine of these were retained because they appeared to provide reliable responses in a different pattern to the majority. These were commonly refractive surgery subjects with postoperative complications. The 16 questionnaires that were discarded contained the same category response for (typically) the last three pages of the questionnaire (these pages contained questions with reversed scales—typically a good QOL would score 1, but for some questions a good QOL scored 5; therefore, a respondent who marked "1" for all the questions on a page suggested responding without reading the questions and therefore unreliable data). This left 316 questionnaires (106 from contact lens wearers, 102 spectacle wearers, and 108 postrefractive surgery).

To ensure the three types of refractive correction have an equal input into the questions retained in the final questionnaire, completed questionnaires were randomly discarded to provide equal numbers of questionnaires from each group. This crucially reduces the potential for bias in favor of one correction type. Random discarding left a final N of 306 (mean age, 27.3 ± 4.9 years), with 102 questionnaire responses from spectacle wearers (mean age, 26.6 ± 5.2 years), 102 contact lens wearers (mean age, 26.9 ± 4.4 years), and 102 refractive surgery subjects (mean age, 28.5 ± 3.9 years). Rasch analysis was then used for item reduction (see Results).

Assessment of the Validity and Reliability of the 20-Item QIRC

The 20-item QIRC questionnaire (Table 1, Appendix) was again administered across several settings across the United Kingdom, as previously described. Questions 14 to 20 regarding feelings of well-being were asked in relation to the subject's refractive correction in that the preamble to the questions included the following text: "We are now interested in the effect that your refractive correction has had on the way you have been feeling. The effect may be obvious (e.g., you may feel that you look better in your new spectacles) or it may be indirect (e.g., you may feel more confident since wearing contact lenses because you feel that you look better)." Three hundred eighty-six questionnaires were returned. Twenty-three questionnaires were discarded because of absent demographic data or >33% missing item responses. Rasch outfit statistics identified 78 possible rogue responders, and after review 42 were retained. This left 327 questionnaires (110 from contact lens wearers, 113 spectacle wearers, and 104 postrefractive surgery). To equalize group sizes, random discarding led to a final N of 312, with 104 questionnaire responses from each refractive correction mode. The validity of the QIRC data was assessed using Rasch analysis, factor analysis, and Cronbach's α, and the test-retest reliability of QIRC was assessed using data from 40 subjects (mean test-retest time, 7.8 ± 7.7 weeks), with some from each refractive correction mode group, using the intraclass correlation coefficient (ICC) and the coefficient of repeatability (COR).[45, 46]

Demographic Data

Ethnicity classification was sourced from the Compendia and Reference section of the National Statistics web site.[43] Socioeconomic classification for the pilot questionnaire was assigned from residential postcode data as the Index of Multiple Deprivation, which is a logarithm comparing each location to the highest socioeconomic status location in Britain (a lower score indicates a higher classification), as sourced from the National Statistics web site.[43] Although this is a recognized method, we were concerned about the accuracy of postcode-based assignment with a small sample such as N = 300. Therefore, we replaced this method for the second phase of the study with a 5-category scale,[47] on which socioeconomic status was classified from self-report of occupation of the primary income earner in the household.[47] This 5-category scale is assumed to be linear; therefore, the mean group scores can be used to compare socioeconomic status.

TABLE 1.
The 20 items included in the QIRC questionnaire, with Rasch fit statistics, item calibration, and factorial validity

	Item Description	Infit mean square (z std)	Outfit mean square (z std)	Item Calibration (SE)	Factor Loadings
1	How much difficulty do you have driving in glare conditions?	1.08 (1.2)	1.18 (2.2)	45.0 (0.8)	0.60
2	During the past month, how often have you experienced your eyes feeling tired or strained?	0.80 (−3.5)	0.87 (−1.8)	49.6 (0.8)	0.46
3	How much trouble is not being able to use off-the-shelf (non prescription) sunglasses?	1.15 (2.0)	1.13 (1.3)	41.3 (0.9)	0.63
4	How much trouble is having to think about your spectacles or contact lenses or your eyes after refractive surgery before doing things; e.g. traveling, sport, going swimming?	0.97 (−0.5)	1.05 (0.7)	45.8 (0.8)	0.55
5	How much trouble is not being able to see when you wake up; e.g. to go to the bathroom, look after a baby, see alarm clock?	0.95 (−0.8)	0.92 (−1.0)	43.8 (0.9)	0.60
6	How much trouble is not being able to see when you are on the beach or swimming in the sea or pool, because you do these activities without spectacles or contact lenses?	1.08 (1.2)	1.06 (0.8)	48.4 (0.9)	0.71
7	How much trouble is your spectacles or contact lenses when you wear them when using a gym/doing keep-fit classes/circuit training etc?	1.03 (0.4)	1.05 (0.5)	39.5 (1.0)	0.59
8	How concerned are you about the initial and ongoing cost to buy your current spectacles/contact lenses/refractive surgery?	1.06 (1.0)	1.04 (0.5)	49.1 (0.8)	0.49
9	How concerned are you about the cost of unscheduled maintenance of your spectacles/contact lenses/refractive surgery; e.g. breakage, loss, new eye problems?	0.85 (−2.6)	0.85 (−2.1)	45.1 (0.8)	0.59
10	How concerned are you about having to increasingly rely on your spectacles or contact lenses since you started to wear them?	0.90 (−1.5)	0.85 (−1.7)	49.9 (0.9)	0.63
11	How concerned are you about your vision not being as good as it could be?	0.94 (−0.9)	0.96 (−0.6)	49.7 (0.8)	0.47
12	How concerned are you about medical complications from your choice of optical correction (spectacles, contact lenses and/or refractive surgery)?	0.98 (−0.2)	0.97 (−0.4)	43.9 (0.8)	0.65
13	How concerned are you about eye protection from ultraviolet (UV) radiation?	1.24 (3.6)	1.32 (3.8)	51.2 (0.8)	0.61
14	During the past month, how much of the time have you felt that you have looked your best?	1.02 (0.2)	0.99 (−0.1)	53.5 (0.7)	0.69
15	During the past month, how much of the time have you felt that you think others see you the way you would like them to (e.g. intelligent, sophisticated, successful, cool, etc)?	0.95 (−0.7)	0.94 (−0.7)	57.0 (0.8)	0.76
16	During the past month, how much of the time have you felt complimented/flattered?	1.19 (2.1)	1.14 (1.4)	62.5 (0.9)	0.66
17	During the past month, how much of the time have you felt confident?	0.85 (−2.1)	0.87 (−1.9)	50.5 (0.7)	0.64
18	During the past month, how much of the time have you felt happy?	0.78 (−3.3)	0.78 (−3.4)	47.5 (0.7)	0.74
19	During the past month, how much of the time have you felt able to do the things you want to do?	1.01 (0.1)	0.99 (−0.1)	39.5 (0.7)	0.40
20	During the past month, how much of the time have you felt eager to try new things?	1.24 (3.1)	1.23 (3.0)	49.1 (0.7)	0.41

RESULTS

90-Item Pilot Questionnaire

The ages of the spectacle and contact lens wearer groups were similar, but the refractive surgery group was slightly older (analysis of variance [ANOVA], $F[2,302] = 9.51$; $p < 0.001$; *post hoc* $p < 0.05$; Table 2). Similarly, the socioeconomic status of the three groups as described by the Index of Multiple Deprivation was similar between spectacle and contact lens wearers, but the refractive surgery group was slightly more affluent ($F[2,302] = 3.44$; $p < 0.05$) and more in line with the UK national average (1.23 ± 0.31; Table 2).[43] Gender and ethnicity distributions were similar between groups (χ^2, $p > 0.05$; Table 2). Ethnicity across the entire cohort was representative of the total UK population,[43] and although the gender distribution shows more female subjects than the total UK population, this is representative of the UK population seeking eye care.[48]

Fig. 1 shows a subject/item map determined by Rasch analysis for the original 90-item QIRC. Subjects (Xs on the left) appear in ascending order of QOL, determined by their average score for all the items, from the bottom of the figure to the top. Items appear on the right of the diagram as item numbers with a decimal representing the response scale boundary. For a 5-category scale, there are four boundaries between categories so that each item is represented in Fig. 1 by four numbers. These are located on the scale in which

TABLE 2.
Demographic characteristics of the sample for the 90-item pilot questionnaire and the 20-item QIRC validation study[a]

	90-item Pilot			20-item QIRC		
	Spectacles	Contact Lenses	Refractive Surgery	Spectacles	Contact Lenses	Refractive Surgery
Age (yr)	26.6 ± 5.5	26.9 ± 4.5	28.5 ± 3.9	24.2 ± 5.9	24.9 ± 5.5	28.7 ± 3.8
Gender (% female)	66	64	62	65	69	73
Socioeconomic status	1.31 ± 0.23	1.30 ± 0.19	1.24 ± 0.26	3.4 ± 0.9	3.5 ± 0.9	3.5 ± 0.8
Race						
% white	86.3	90.2	84.3	93.3	92.3	90.4
% Asian	6.9	1.0	9.8	1.0	6.7	1.9
% black	1.0	1.0	1.0	4.8	0.0	1.0
% mixed	3.0	3.9	0.0	0.0	2.9	1.0
% other	2.0	0.0	0.0	0.0	0.0	1.9

[a] Socioeconomic status was determined by the National Statistics log of the index of multiple deprivation[43] (for the subject's postcode) in the pilot study and by using a five-category Occupational Classification[47] (for the household income earner) for the 20-item QIRC validation study. The three groups for the 90-item pilot were similar on all measures except that the refractive surgery group was slightly older (analysis of variance $F2,302 = 9.51$; $p < 0.001$; post hoc $p < 0.05$) and more affluent ($F2,302 = 3.44$; $p < 0.05$; in line with national averages). The three groups for the 20-item QIRC were similar on all measures except that the refractive surgery group was slightly older (analysis of variance, $F2,301 = 19.33$; $p < 0.001$; post hoc $p < 0.001$).

the most likely response changes from one category to the next. Items appear in ascending order of severity of impact on refractive error-related QOL from the bottom of the figure to the top. Subjects and items appear along the same scale, which is a linear transformation of the Rasch logit scale to fit a 0 to 100 scale (Winsteps $U_{mean} = 50.73$; $U_{scale} = 6.89$). For this dataset, many items (especially functioning [items 1 to 19] and symptoms [items 20 to 35]) are irrelevant to the QOL of the patients, which is seen as the X's located higher and item numbers located lower. This illustrates inadequate targeting of item difficulty to patient ability. If the items were well targeted to the subjects, the means of the two distributions, denoted in Fig. 1 by M, would be close to each other. We attempted to improve targeting of items to subjects through response scale reduction and item reduction.

No item had >6% of responses in the "extreme" end category, suggesting that a 5-category response scale contained too many categories. Combining response 1 "extreme" and response 2 "quite a lot" reduced the underutilization of the end category and improved targeting of items to subject QOL,[17, 24] as illustrated by a reduction in the difference between the mean value for the subjects and the mean value of the items from 11.30 units to 9.68 units. There was still a significant mismatch of item targeting to subjects. The combined response category was still underused for many items, although not for the well-being items. For the 74 non–well-being items, the combined end category was still empty for 13 items and held <10% of the responses for 66 items. Therefore, it was decided to also combine the response category "a moderate amount" with the already combined "quite a lot" and "extreme." This further reduced the difference between the mean value for the subjects and the mean value of the items to 8.53 units. Shortening the response scale did not alter subject separation (5-category = 3.21; 4-category = 3.26; 3-category with 4-category for well-being items = 3.30). Subject separation is an indication of the precision with which the variability present in the people is captured by the test. This is expressed as the ratio of the adjusted SD to the root-mean-square error. Therefore, a higher subject separation indicates subjects being significantly different in ability across the measurement distribution.

Rasch analysis was then used to remove poorly fitting items from the questionnaire. Item reduction was an iterative process, with one item removed at a time and the fit to the model re-estimated accordingly (fit is relative; therefore, removal of items will lead to changes in fit). Set criteria were used to identify candidate items for removal,[17] and these, in order of priority, were:

Infit mean square outside 0.80 to 1.20
Outfit mean square outside 0.70 to 1.30
Item with mean furthest from subject mean (see Fig. 1)
High proportion of missing data (>50%)
Ceiling effect: A high proportion in item end-response category (>50%)
Skew and kurtosis outside −2.00 to +2.00

The item with the highest number of candidate criteria, ordered by priority, was removed first. As mentioned earlier, items with low infit and outfit may be redundant or noncontributory because they lack variance, and high item infits or outfits suggest that the items measure something different to the overall scale.[44] A high item outfit may also indicate that an item is affected in some patients but not in all. Such items would be acceptable as long as they are not too extreme. Based on this rationale, we determined that infit mean square should drive item reduction; therefore, more stringent criteria were used for infit, and more lenient criteria were used for outfit.

Seventy items were removed in total. This reduced the questionnaire to 20 items (Table 1), ensuring a low respondent burden. This figure was reached when all the items provided good infit and outfit values, with no significant missing data or ceiling effect and an acceptable patient separation (>2.00). Reducing the number of items further led to decreased patient separation. The person/item map for the 20-item questionnaire, with a much improved targeting of item QOL to subject QOL, is shown in Fig. 2. This again is fit to a

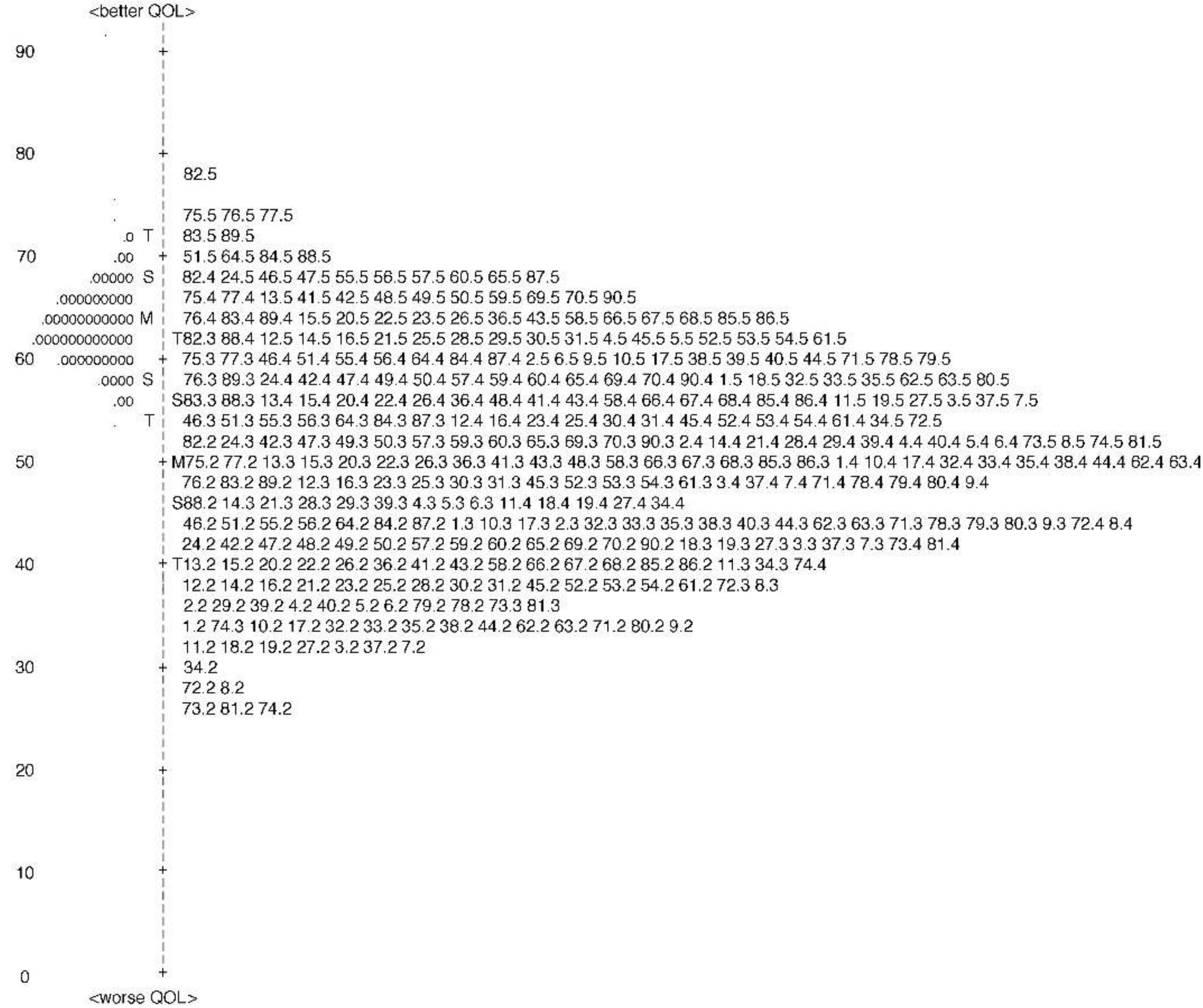

FIGURE 1.

Person/item map for the 90-item pilot Quality of Life Impact of Refractive Correction (QIRC) questionnaire with the cross-over point for each response category for each item shown on the right. On the left of the dashed line are the subjects, represented by ○ (○ = 5 subjects). On the right are the cross-over points between each item category (point of the scale where the answer category is most likely to be chosen by a subject with that quality of life [QOL]). Subjects with poorer QOL are near the bottom of the diagram, and subjects with better QOL are near the top. Items that are unaffected in patients with refractive correction are near the bottom of the diagram, and items near the top are those strongly affected by refractive corrections. The scale is in units (0 to 100), and abbreviations on the diagram are M = mean, S = 1 SD from the mean, and T = 2 SD's from the mean.

0 to 100 scale (Winsteps $U_{mean} = 47.83$; $U_{scale} = 9.69$). The difference in means between the subject and items was reduced to 2.45 units, and acceptable person separation (2.03) was retained.

Assessment of the Validity and Reliability of the 20-Item QIRC

The three groups were different for age (ANOVA, $F[2,301] = 19.33$; $p < 0.001$) because the refractive surgery group was significantly older ($p < 0.001$) than the spectacle and contact lens groups, which were similar to each other ($p > 0.05$). The three groups were similar for gender ($\chi^2 > 0.05$), ethnicity ($\chi^2 > 0.05$), and socioeconomic status ($\chi^2 > 0.05$). Socioeconomic status and ethnicity were similar to the total UK population,[43, 47] and gender was similar to the UK population seeking eye care.[48] The items (mean score, 50.3 ± 7.3) were well targeted to the subjects (mean score, 47.8 ± 5.5) with a mean difference of 2.45 (scale range, 0 to 100) units. Rasch analysis yields valid model statistics (person separation, 2.03; reliability, 0.80; root-mean-square measurement error, 3.25; mean square ± SD infit, 0.99 ± 0.38; outfit, 1.00 ± 0.39), and fit statistics show all the items fit within a range of infit from 0.78 to 1.24 and for outfit 0.78 to 1.32. Thus, the variance within items extends from 22% less than the expected to 24% (for infit) and 32% (for outfit) more than the expected. Unrotated factor analysis established a principal factor with loadings from

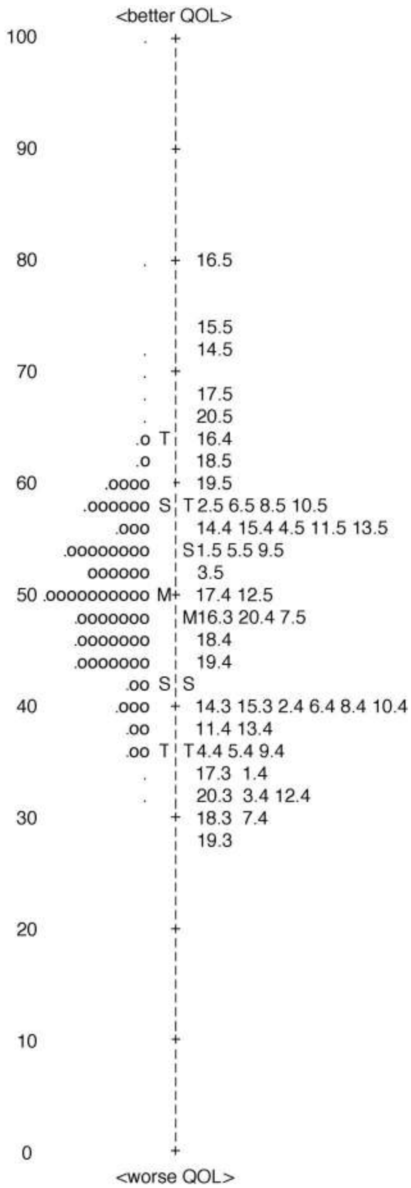

FIGURE 2.

Person/item map for the 20-item final Quality of Life Impact of Refractive Correction (QIRC) questionnaire with the cross-over point for each response category for each item shown on the right. On the left of the dashed line are the subjects, represented by ○ (○ = 4 subjects). On the right are the cross-over points between each item category (point of the scale where the answer category is most likely to be chosen by a subject with that quality of life). This question group exhibits excellent targeting of items to subjects.

0.40 to 0.76 (Table 1), and Cronbach's α for the scale was 0.78. These three findings suggest all the items measure a unitary concept, without redundancy (no loadings, >0.80). There were no significant differences between test and retest QIRC scores (mean, −0.41 units; $F[1,39] = 0.543$; $p > 0.05$), and test-retest reliability was good, with an ICC of 0.88 and a COR of ± 6.85 units.

DISCUSSION

The QIRC questionnaire was rigorously developed using conventional techniques and Rasch analysis to ensure construct validity. Importantly, this included thorough domain and item selection using a broad base of contributors, including focus groups. The 647 items were carefully reduced using focus groups and a pilot questionnaire.

The item reduction phase highlights an advantage of Rasch analysis beyond the obvious importance that it provides a quantitative score that is a valid linear measurement. The pilot questionnaire, although influenced by suggestions from lay people, was principally clinician driven and contained many questions relating to visual function (21%) and symptoms (18%). These are similar proportions to those found for the RSVP (about 23% visual function and 33% symptoms)[4] and NEI-RQL (about 31% visual function and 36% symptoms).[9] However, Rasch analysis clearly indicated that subjects with corrected refractive error have relatively few problems with visual function and had few symptoms (Fig. 1); therefore, many of these questions were not required with only one of each being retained in the 20-item QIRC. These items did not misfit but tended to have high floor effects and were therefore noncontributory. If they had been left in, QIRC would have poorly targeted the population.

This may be why other vision-related QOL instruments that principally contain such items, such as the RSVP, lack sensitivity in subjects with refractive error.[6] Similarly, the NEI-RQL has some ceiling effect problems—54% for the "Activity Limitations" subscale and 40% on the "Glare" subscale.[3] This likely arises from clinicians overemphasizing the importance of visual function losses and symptoms because they are most prevalent when a patient is examined by a clinician (i.e., presenting complaint). It would appear that subjects at large have few symptoms or problems with visual function that are not corrected and/or that they view such problems as minor (perhaps because they assume them to be easily resolved), and issues such as convenience, cost, health concerns, and appearance determine the influence of refractive error correction on QOL (Table 1). Rasch analysis was essential to providing this insight into item relevance or targeting. However, this is population dependent, and QIRC was developed to find an ideal item set to suit all three modes of refractive correction. It may be that the ideal item set would be different if the questionnaire were developed for spectacle wearers alone or contact lens wearers alone. Although this will be the subject of future analyses, herein we present a questionnaire that will suit spectacle, contact lens, and refractive surgery-corrected subjects.

The Rasch model and fit statistics for each item, factorial validity, and Cronbach's α demonstrate that QIRC is reliable and valid (Table 1, Fig. 2). Cronbach's α has been used as a reliability coefficient in some reports but really just represents the unidimensionality of a questionnaire because it is essentially determined by the average of the correlation coefficients between items; therefore, exceptionally high values of Cronbach's α (>0.90) may indicate redundancy.[49] Redundancy is a problem if the process of creating the "overall score" for the questionnaire involves just adding all the item scores together. In such a case, the overall score over-weights the importance of the issue that is served by redundant items.[50] Similarly, Cronbach's α is not independent of the number of items and may be elevated by including many items. For these reasons, Cronbach's α should probably be considered to be more of a traditional measure than a useful measure.[16] Factorial validity is a more useful indicator of internal consistency. In the final QIRC, all the items had a factor loading between 0.40 and 0.76, which indicates significant but not redundant correlation. This confirms the Rasch model fit findings that all the items contribute significantly to the overall measure and that all measure a related concept. We suggest this concept is QOL related to their correction of refractive error.

Retesting on stable patients within an average period of 2 months demonstrated good repeatability with both methods, ICC (0.88) and Bland-Altman analysis (COR, ±6.85). The responsiveness of QIRC to the provision of various refractive corrections will be examined in a subsequent study.

The 90-item and 20-item questionnaires were implemented on samples that approximated the UK population demographics for refractive error correction in terms of age, gender, socioeconomic classification, and ethnicity. Although the groups were similar to population norms, the refractive surgery groups for both sets of data collection were older than the spectacle and contact lens wearers. For the 90-item questionnaire, the refractive surgery group also appeared to be more affluent. This is probably inevitable because it reflects UK population averages for uptake of refractive surgery in the prepresbyope.[51] This is important because it shows the population was representative of the United Kingdom and does not pose a problem because no comparison between groups has been made.

Care was also taken to discard poor data. We deliberately worded several questions to allow reversal of the scale direction to catch careless responders. Rasch analysis provides a powerful test to detect such rogue data through the outfit statistic. Thus, we were able to prevent poor data contaminating the dataset. The dataset was equalized in numbers of questionnaires completed by wearers of each correction mode. This was valid because it was done randomly and important because it prevents any one correction type from having greater influence on the composition of the final QIRC questionnaire.

Limitations of QIRC include that it has only been developed for the prepresbyopic population. However, this was intentional; we found during item selection that many presbyopic-specific items were suggested by the professional focus groups. These included issues related to needing to remove reading glasses to see in the distance, problems with bifocal segments, and multifocal distortions. Because these items would not be applicable to prepresbyopes, this might cause problems with model fit, targeting to subjects and measurement integrity. Although this has not been tested experimentally, it is likely that a single questionnaire, which tries to suit both groups, would be suboptimal. Nevertheless, QIRC could be used in presbyopic subjects; it just would not tap presbyopic-specific issues. Validation of QIRC in presbyopic subjects is the topic of a study in progress.

Another possible limitation is the small number of items, 20. Although this makes for a test of low burden, it raises the possibility that an issue applicable to only a small number of people may not be included, and therefore its impact on QOL may not be measured. Certainly, this is possible. However, for the most part, responses to items that target the same underlying issue will behave in the same way; as long as the questionnaire has items that are representative of the issues, accurate measurement can be made. For example, if a contact lens wearer is experiencing red eye, this could be measured with a question specifically asking about red eyes; however, if no such question is included, it does not mean that the impact of red eyes on QOL is missed. This issue may be tapped by questions on appearance, sore/tired eyes, concerns about ocular health, or convenience (e.g., thinking about eyes). Although this is likely how QIRC functions for most issues, it is possible some issues are untapped.

In conclusion, we present the QIRC questionnaire. This is a 20-item questionnaire reporting a single-valued score of QOL in the refractive corrected. This instrument has several advantages over existing instruments: developed on equal numbers of subjects corrected by spectacles, contact lenses, and refractive surgery; proven with Rasch analysis that all the items measure a single content area; and scaled using Rasch analysis to be a true linear measurement of QOL in which items are weighted for their impact on QOL.

ACKNOWLEDGMENTS

We thank Mike Stevens, Helen Scott, Jenny Burr, Simon Gelsthorpe, and Alison Torn for assistance with professional and lay focus groups and Dr. Mike Linacre for help with various aspects of Rasch analysis.

Supported by a grant from Johnson & Johnson Vision Care Europe. NHMRC Sir Neil Hamilton Fairley Fellowship 007161 supports K. Pesudovs. Ultralase provided access to refractive surgery subjects.

Received February 10, 2004; accepted June 16, 2004.

APPENDIX

Both the questionnaire and score sheet are available online only at www.optvissci.com.

REFERENCES

1. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ 2002;324:1417.
2. Vision Watch. Vision Correction Market Review. New York: Jobson Publishing, 2003.
3. 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;110:2292–301.
4. Schein OD. The measurement of patient-reported outcomes of refractive surgery: the refractive status and vision profile. Trans Am Ophthalmol Soc 2000;98:439–69.
5. Walline JJ, Bailey MD, Zadnik K. Vision-specific quality of life and modes of refractive error correction. Optom Vis Sci 2000;77: 648–52.
6. Nichols JJ, Mitchell GL, Zadnik K. The performance of the refractive status and vision profile survey in a contact lens clinical trial. Ophthalmology 2001;108:1160–6.
7. McDonnell PJ, Lee P, Spritzer K, Lindblad AS, Hays RD. Associations of presbyopia with vision-targeted health-related quality of life. Arch Ophthalmol 2003;121:1577–81.
8. McDonnell PJ, Mangione C, Lee P, Lindblad AS, Spritzer KL, Berry S, Hays RD. Responsiveness of the National Eye Institute Refractive Error Quality of Life instrument to surgical correction of refractive error. Ophthalmology 2003;110:2302–9.
9. Berry S, Mangione CM, Lindblad AS, McDonnell PJ. Development of the National Eye Institute refractive error correction quality of life questionnaire: focus groups. Ophthalmology 2003;110: 2285–91.

10. Nichols JJ, Mitchell GL, Saracino M, Zadnik K. Reliability and validity of refractive error-specific quality-of-life instruments. Arch Ophthalmol 2003;121:1289–96.
11. 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.
12. Rose K, Harper R, Tromans C, Waterman C, Goldberg D, Haggerty C, Tullo A. Quality of life in myopia. Br J Ophthalmol 2000;84: 1031–4.
13. Hill JC. An informal satisfaction survey of 200 patients after laser in situ keratomileusis. J Refract Surg 2002;18:454–9.
14. Day H, Jutai J. Measuring the psychosocial impact of assistive devices: the PIADS. Can J Rehab 1996;6:135–41.
15. Likert R. A technique for the measurement of attitudes. Arch Psychol 1932;140:1–55.
16. Massof RW. The measurement of vision disability. Optom Vis Sci 2002;79:516–52.
17. 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.
18. 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.
19. 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.
20. White LJ, Velozo CA. The use of Rasch measurement to improve the Oswestry classification scheme. Arch Phys Med Rehabil 2002;83: 822–31.
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:I25–36.
22. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford: Oxford University Press, 1995.
23. Wright BD, Masters GN. Rating Scale Analysis. Chicago: MESA Press, 1982.
24. Linacre JM. Investigating rating scale category utility. J Outcome Meas 1999;3:103–22.
25. Linacre JM. A User's Guide to Winsteps: Rasch-Model Computer Program. Chicago: Mesa Press, 2002.
26. Wright BD, Stone MH. Best Test Design. Chicago: MESA Press, 1979.
27. Fallowfield L. The Quality of Life: The Missing Measurement in Health Care. London: Souvenir Press, 1990.
28. Bowling A. What things are important in people's lives? A survey of the public's judgements to inform scales of health related quality of life. Soc Sci Med 1995;41:1447–62.
29. Felce D. Defining and applying the concept of quality of life. J Intellect Disabil Res 1997;41:126–35.
30. Seed P, Lloyd G. Quality of Life. London: Jessica Kingsley, 1997.
31. Oliver JPJ. Quality of Life and Mental Health Services. London: Routledge, 1996.
32. Margolis MK, Coyne K, Kennedy-Martin T, Baker T, Schein O, Revicki DA. Vision-specific instruments for the assessment of health-related quality of life and visual functioning: a literature review. Pharmacoeconomics 2002;20:791–812.
33. Massof RW, Rubin GS. Visual function assessment questionnaires. Surv Ophthalmol 2001;45:531–48.
34. Schein OD, Vitale S, Cassard SD, Steinberg EP. Patient outcomes of refractive surgery. The refractive status and vision profile. J Cataract Refract Surg 2001;27:665–73.
35. Vitale S, Schein OD, Steinberg EP, Ware JE Jr. The refractive status and vision profile (RSVP): initial psychometric attributes. Invest Ophthalmol Vis Sci 1997;38:S841.
36. Vitale S, Schein OD, Meinert CL, Steinberg EP. The refractive status and vision profile: a questionnaire to measure vision-related quality of life in persons with refractive error. Ophthalmology 2000;107:1529–39.
37. Demers L, Monette M, Lapierre Y, Arnold DL, Wolfson C. Reliability, validity, and applicability of the Quebec User Evaluation of Satisfaction with assistive Technology (QUEST 2.0) for adults with multiple sclerosis. Disabil Rehabil 2002;24:21–30.
38. Day H, Jutai J, Campbell KA. Development of a scale to measure the psychosocial impact of assistive devices: lessons learned and the road ahead. Disabil Rehabil 2002;24:31–7.
39. Fitzpatrick R, Jenkinson C, Klassen A, Goodacre T. Methods of assessing health-related quality of life and outcomes for plastic surgery. Br J Plast Surg 1999;52:251–5.
40. Alsarraf R. Outcomes research in facial plastic surgery: a review and new directions. Aesthetic Plast Surg 2000;24:192–7.
41. Nagata C, Ido M, Shimizu H, Misao A, Matsuura H. Choice of response scale for health measurement: comparison of 4, 5, and 7-point scales and visual analog scale. J Epidemiol 1996;6:192–7.
42. Skevington SM, Tucker C. Designing response scales for cross-cultural use in health care: data from the development of the UK WHOQOL. Br J Med Psychol 1999;72:51–61.
43. Office for National Statistics. National Statistics Online [National Statistics Web site]. 2001. Available at: http://www.statistics.gov.uk/default.asp. Accessed February 16, 2001.
44. Smith RM. Person fit in the Rasch model. Educ Psychol Measurement 1986;46:359–72.
45. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.
46. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1: 307–10.
47. The Market Research Society. Occupation Groupings. A Job Dictionary. London: The Market Research Society, 2003.
48. Pointer JS. An optometric population is not the same as the general population. Optometry Prac 2000;1:92–6.
49. DeVellis RF. Scale Development: Theory and Applications. Newbury Park, CA: Sage, 1991.
50. Bland JM, Altman DG. Statistics notes: validating scales and indexes. BMJ 2002;324:606–7.
51. Mintel Group. Optical Goods and Eyecare. London: Mintel Group, 2002.

Konrad Pesudovs
Department of Optometry
University of Bradford
Richmond Road
Bradford, West Yorkshire, BD7 1DP
United Kingdom
e-mail: K.Pesudovs@bradford.ac.uk