Catquest-9SF patient outcomes questionnaire

Nine-item short-form Rasch-scaled revision of the Catquest questionnaire

Mats Lundström, MD, PhD, Konrad Pesudovs, PhD

PURPOSE: To assess and optimize the Catquest questionnaire for measuring patient-reported outcomes of cataract surgery using Rasch analysis.

SETTING: Fifty-eight ophthalmic surgical units in Sweden.

METHODS: Catquest questionnaires (n = 21 364) from the Swedish National Cataract Register were selected and randomized to 2 groups. Data from 10 486 questionnaires were comprehensively Rasch analyzed using a 4-Andrich rating scale model in Winsteps software. A revised version of Catquest was developed (Catquest-9SF) and tested in 10 886 patients for validity and responsiveness to cataract surgery.

RESULTS: Only the visual disability subscale formed a valid measurement scale. This could be enhanced through the addition of the 2 global assessment items; however, the symptoms and frequency of performing the activities items did not contribute to the measurement. The 9-item short-form version (Catquest-9SF) had ordered response thresholds and good person separation (2.65) and was largely free from differential item functioning. All items fit a single overall construct (infit range, 0.75 to 1.29; outfit range, 0.70 to 1.39) and unidimensional by principal components analysis. The items were well targeted to the preoperative participants (0.34 logit difference in means). The score correlated with visual acuity ($r = 0.43$ preoperatively; $r = 0.48$ postoperatively) and was highly responsive to cataract surgery (preoperatively -0.32 ± 2.15 logits; postoperatively -3.21 ± 2.50 logits ($P<.0001$).

CONCLUSIONS: The 9-item Rasch-scaled Catquest-9SF was highly valid in measuring visual disability outcomes of cataract surgery. Its brevity makes it suited to routine clinical use, and a raw-data to Rasch-measure conversion simplifies application.

J Cataract Refract Surg 2009; 35:504–513 © 2009 ASCRS and ESCRS

Many visual disability questionnaires have been validated in cataract patients; these include the Activities of Daily Vision Scale (ADVS),[1] the Visual Disability Assessment (VDA), and the Visual Functioning 14 (VF-14)[2]. The questionnaires have been shown to be sensitive to clinically meaningful change after cataract surgery.[3,4] However, further validation using item-response theory, in particular Rasch analysis, has shown limitations in questionnaire development and validation not previously highlighted using traditional classic test theory.[5–7]

Another frequently used visual disability questionnaire is the Catquest.[8] In 1995, the National Swedish Cataract Register began collecting data on patients' self-assessed visual functions.[9] The Catquest was used to measure changes in patient-reported visual function 6 months after cataract surgery compared with before surgery.[10] The design of the questionnaire has been described.[8] Basically, questions were asked in 4 areas. The outcome was evaluated in each area by comparing a score before surgery and after surgery. The total outcome of surgery was evaluated using a decision tree, which means that the more areas that improved after surgery, the better the outcome. However, in each area of the Catquest, the basic principle was to add scores achieved from patients' choice of response options.

This summary scoring, also termed Likert scoring, allocates an ordinal numerical value to a participant's response for each item. However, this method of scoring has limitations, not least of which is the erroneous assumption that this method of scoring produces an

Published by Elsevier Inc.
0886-3350/09/$—see front matter
doi:10.1016/j.jcrs.2008.11.038

interval scale. Modern test theory, including Rasch analysis, has shown the invalidity of summary scoring[11–13] and resolves the problem by providing linear interval transformation of the ordinal raw score, thereby permitting the use of parametric statistical techniques on the questionnaire data.[11,14,15] Other unique evaluations available in Rasch analysis include how well item difficulty targets person ability in the population being assessed and scale validity assessment, in particular item and person fit to the overall construct.[6,13,16] As mentioned above, Rasch analysis has been extensively used to review and improve existing questionnaires that were constructed using Likert scales.[5–7,15,17] The Rasch analysis technique has also been used to construct new questionnaires.[18,19]

In this study, we applied Rasch analysis to a database of Catquest questionnaires completed before and after cataract surgery.[7] The purpose was to assess and reengineer the Catquest questionnaire using Rasch analysis to optimize item fit to the construct, minimize test length, and create a linear measure of visual disability for measuring the outcomes of cataract surgery.

PATIENTS AND METHODS

Catquest

The Catquest questionnaire contains questions for evaluating the benefit of cataract surgery.[8] There are questions within 4 content areas: frequency of performing activities (6 questions), perceived difficulty in performing daily-life activities (7 questions), global questions about difficulties in general and satisfaction with vision (2 questions), and cataract symptoms (2 questions). There are 4 (summary scoring value) response options for the perceived difficulty levels as follows: 4 = very great difficulty; 3 = great difficulty; 2 = some difficulty; 1 = no difficulty. Therefore, a lower score is better and a higher score is worse. The 2 items on cataract symptoms also have these 4 response options. For satisfaction with vision, the 4 response options are as follows: 4 = very dissatisfied; 3 = rather dissatisfied; 2 = fairly satisfied; 1 = very satisfied. The frequency of performing the activity items have 4 response options: 4 = do not do the activity; 3 = do the activity rarely (often once a week); 2 = do the activity more frequently (2 to 4 times a week or for television watching, at least 1 hour per day); 1 = do the activity frequently (every day or for television watching, several hours per day). There are also questions about other things such as home help, other diseases, and car driving/employment. These latter items have been used as demographic variables, not to evaluate the benefit of surgery, and were therefore excluded from this analysis. The items are presented in the same format in both the preoperative and postoperative versions of the questionnaire. In this study, both preoperative and postoperative were included to allow evaluation of the validity of the questionnaire in both situations.

The Catquest has been used in the Swedish National Cataract Register since 1995.[9] The participating surgical units have, on a voluntary basis during 1 month each year, used the questionnaire on all patients having cataract surgery during that month. The number of participating units has varied between 25 and 35. In the database, there are 23614 participants with a completed questionnaire before and after surgery in the period from 1995 to 2006. The Catquest has been used in several studies in subgroups of cataract patients.[20–23] All questionnaires were completed in Swedish; therefore, the information about the question and response formats presented here represents a translation.

Patients

In this study, data from 1995 to 2005 from 58 surgical units in Sweden were used. Before the Swedish National Cataract Register began collecting data on patients' self-assessed visual function, the questionnaire (Catquest) and the method was approved by an ethics committee according to the Declaration of Helsinki and by the Swedish Data Inspection Board. The patients were informed about the study according to Swedish law. The data were split into 2 groups of approximately equal size. The division was made by random selection of the Statistical Package for the Social Sciences software (version 15.0, SPSS, Inc.). The first group was used for the assessment and redevelopment of the Catquest (development group). The second group was used as an independent population to test the validity of the revised Catquest and the outcomes of cataract surgery (validation group).

Rasch Analysis

The Catquest data were assessed for fit to the Rasch model[24] using Winsteps software (version 3.63.2, Linacre[25,26]) and the Andrich[27] version of Rasch model estimates based on joint maximum likelihood estimation. An individual Andrich rating scale was applied for each question format. Activity level (6 questions), perceived difficulties in performing daily-life activities (7 questions), global questions about difficulties in general and satisfaction with vision (2 questions), and cataract symptoms (2 questions) all have a different format. Therefore, the analysis used a 4-Andrich rating scale design. Each rating scale was analyzed separately; different combinations of the scales were included in a single analysis to determine whether a more comprehensive overall measure was possible. The Catquest questionnaire should be valid for measurement for both preoperative and postoperative patient data. Therefore, Rasch analysis was performed on preoperative and postoperative data stacked as a single data set (10478 cases but 20956 response sets).[28]

Submitted: September 19, 2008.
Final revision submitted: November 10, 2008.
Accepted: November 12, 2008.

From the EyeNet Sweden (Lundström), Blekinge Hospital, Karlskrona, Sweden, and the NH&MRC Centre for Clinical Eye Research (Pesudovs), Flinders University and Flinders Medical Centre, Flinders, Australia.

Neither author has a financial or proprietary interest in any material or method mentioned.

Corresponding author: Mats Lundström, MD, PhD, EyeNet Sweden, Blekinge Hospital, SE-371 85 Karlskrona, Sweden. E-mail: mats.lundstrom@ltblekinge.se.

Rasch analysis assumes that the probability of a respondent affirming an item is a logistic function of the relative distance between the item location and the respondent location on a linear scale. Hence, it is anticipated that the probability of endorsing a particular category will increase monotonically with the difference between the respondent's level of difficulty in performing daily activities and the level of difficulty required for the task. When the data meet the Rasch model expectations, a transformation of the ordinal raw score into a true Rasch scale is achieved.[25,27] In the case of the Catquest, a positive person logit score suggests that the person's level of ability is lower than the mean required level of difficulty for the items. Conversely, if a person logit score is negative, the person's perceived level of ability is higher than the average required level of difficulty.

The key indicators of overall scale performance are person separation and person separation reliability. These measures are related and indicate how well the items of the instrument separate the respondents. Reliability ranges between 0 and 1, with larger values indicating a greater ability to distinguish between the strata of person ability. Separation should be at least 2.00, and separation reliability should be at least 0.80, indicating 3 strata can be discriminated, as a minimum level of performance to constitute a valid measure.

The presence of disordered response category thresholds was examined before overall fit of the data to the model was assessed. Disordered thresholds occur when participants have difficulty discriminating between ordered response options, with categories expected to be "harder" being endorsed as "easier," or alternatively may represent category interchangeability rather than true disordering. Threshold ordering is essential for the calculation of person and item calibrations, so disordering needs to be resolved. Interchangeable categories can be combined into a single category to ensure ordered thresholds.

Overall fit of the data to the model was assessed using 2 overall-fit statistics: infit mean square and outfit mean square (sum of the squared standardized residuals). Both infit and outfit mean squares have an expected value of 1. Values less than 0.70 represent items that overfit the model and are too predictable; that is, they have at least 30% less variation than expected. Conversely, mean squares greater than 1.30 represent misfit with at least 30% more variance than would be expected and suggests that the item measures something different than the overall scale.[29]

Unidimensionality provides further evidence that the instrument is measuring the underlying trait (visual disability). In addition to item-fit statistics, unidimensionality was assessed using principal components analysis (PCA) of the residuals and was formally tested in Winsteps by 3 indications. The first indication is comparison of the amount of variance explained empirically and by the model; multidimensionality elevates the variance explained by the model, thus making it appear greater than the variance explained empirically.[25] The second indication is the amount of variance explained by the first contrast (additional dimension); while this can be tested for significance,[29] it is not appropriate with this large dataset.[30] A threshold of 1.5 Eigenvalue is a suitable, albeit strict, definition of multidimensionality.[31,32] The third indication is an examination of the pattern of factor loadings on the first component to determine "subsets" of items ("positive" and "negative" loadings subsets).

Misfit of the data to the Rasch model could occur because of differential item functioning (DIF), in which different groups of the sample (ie, based on sex, comorbidity, cataract status, and so forth) respond differently to individual item(s) despite having equal levels of the underlying trait. Inspection of the raw differences in item calibration between groups was used to identify DIF. A shift less than 0.50 logits was considered no DIF; a shift of 0.50 to 1.0 logits, minimal but probably inconsequential DIF; and a shift greater than 1.0 logits, at the threshold for notable DIF.[33,34] Testing for DIF was performed between groups of sex, age (<65, 65 to 85, >85 years), with and without ocular comorbidity, first-eye and second-eye cataract surgery, and before surgery and after surgery.

Targeting of the items to the population was also assessed to determine whether the questionnaire items were appropriate for people with cataract. Poorly targeted instruments are often limited by floor or ceiling effects, which are displayed by uneven spread of items across the full range of respondent's scores and/or insufficient items to assess the full range of the sample trait. Targeting is assessed by the pattern of the distributions appearing on a person–item map and by the difference in the value of the person and item mean scores.[18]

Validation Phase

Construct validation involved the testing of 2 hypotheses. The first was that visual disability would correlate with visual acuity. This was tested using the visual acuity data from the registry and Pearson correlation coefficients (r). The second hypothesis was that the visual disability score would improve after cataract surgery. The item–category thresholds determined during the development phase were applied to the person responses of the second group of participants. Overall scores were calculated for preoperative and postoperative data to determine the impact of cataract surgery on the Catquest score. These data were then used to calculate the effect size (the difference between the preoperative score and postoperative score divided by the preoperative standard deviation).[35]

RESULTS

Patients

Between 1995 and 2005, 21 364 Catquest questionnaires from 58 surgical units in Sweden were completed. The mean age of the patients in the database was 75.9 years ± 9.57 (SD), and 66.1% were women. Table 1 shows the characteristics of the patients in the studied database. The development group comprised 10 478 patients and the validation group, 10 886 patients. There were no significantly differences between groups in any characteristic ($P<.01$).

Difficulty in Performing 7 Daily-Life Activities

The response scale thresholds were ordered across the 7 items. Category probability curves for the Andrich rating scale for these items are shown in Figure 1. The overall fit of the data to the model seemed good, indicating that this area formed a valid measure. The real person separation was 2.18 (model person separation 2.41), the person separation reliability, 0.83; the infit range, 0.81 to 1.24 (mean 1.0); and the outfit range,

Table 1. Characteristics of the 2 patient samples randomly selected from the Swedish National Cataract Register.

Characteristic	Development Group	Validation Group
Patients (n)	10 478	10 886
Mean age (y) ± SD	75.87 ± 9.60	75.94 ± 9.54
Sex, n (%)		
Female	6948 (66.3)	7166 (65.8)
Male	3529 (33.7)	3720 (34.2)
First-eye surgery (%)	62.8	62.6
Second-eye surgery (%)	37.2	37.4
Sight-threatening ocular comorbidity, n (%)	3478 (33.2)	3680 (33.8)
Best corrected visual acuity		
Before surgery		
Mean logMAR ± SD	0.59 ± 0.30	0.59 ± 0.30
Mean Snellen	6/24	6/24
Fellow eye		
Mean logMAR ± SD	0.27 ± 0.28	0.27 ± 0.28
Mean Snellen	6/12	6/12
After surgery		
Mean logMAR ± SD	0.15 ± 0.23	0.15 ± 0.22
Mean Snellen	6/7.5	6/7.5

0.82 to 1.24 (mean 0.98). The mean of patients and items were reasonably matched, with a mean difference of −1.18 logit (0.50 preoperatively and −2.15 postoperatively). Cronbach's α was 0.91.

Frequency of 6 Daily-Life Activities

The frequency items did not form a valid measure. The real person separation was 1.06 with a person separation reliability of only 0.53; therefore, this set of items was unable to discriminate the patients. However, the response scale thresholds were disordered, with category 3 (do the activity rarely) never being the threshold most likely chosen and chiefly falling under the range of category 4 (do not do the activity). Therefore, category 4 was combined with category 3. However, this did not improve model performance; the real person separation was 0.92 with a person separation reliability of only 0.46.

Cataract Symptoms

With only 2 items, the symptoms scale did not form a valid scale. The real person separation was 0.00 with a person separation reliability of only 0.00.

Global Assessment

With only 2 items, the global assessment scale also did not form a valid scale. The real person separation was 0.44 with a person separation reliability of only 0.16.

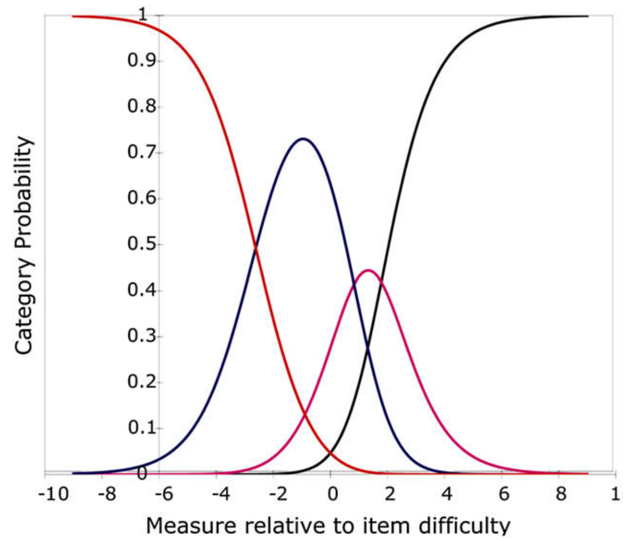

Figure 1. Category probability curves for the 7 visual disability items.

Combinations of Scales

If all items were included in a 4-rating scale analysis, a valid overall scale was formed. The real patient separation was 2.46 and the patient reliability, 0.86. The infit range was 0.55 to 1.92 and the outfit range, 0.53 to 2.41. The symptoms and frequencies items fit poorly. If the symptoms items were retained and the frequency items were removed, a valid overall scale was formed; person separation was 2.53 and reliability, 0.86. However, the symptoms items still grossly misfit (infit range 1.84 to 2.25; outfit range 1.80 to 2.44). Removal of the symptoms items and reinstatement of the frequency items again produced a valid measure; real patient separation was 2.48 and patient reliability, 0.86. However, the frequency items fit poorly (infit range 0.95 to 2.00; 1.04 to 2.88). Removal of the frequency items improved the overall scale performance; real patient separation was 2.65 and patient reliability, 0.88. All items then fit a single overall construct (infit range 0.75 to 1.29; outfit range 0.70 to 1.39). Effectively, this means the 2 global items fit very well with the measurement of visual disability, creating a more reliable measurement scale than the 7 visual disability items alone.

Catquest-9SF

Based on the findings of the analyses combining scales, the 7 disability items were put together with the 2 global items to create Catquest-9SF, a 9-item short-form (SF) measure (Table 2). The response categories were ordered (Figures 1 to 3). All items fit a single overall construct; the infit and outfit for each item are shown in Table 2. Further evidence of

Table 2. The Catquest-9SF questionnaire with item difficulty calibration, infit and outfit mean square, standardized fit statistics, and preoperative versus postoperative DIF.

Item	Item Calibration (Standard Error)	Infit		Outfit		DIF Preop to Postop
		MNSQ	ZSTD	MNSQ	ZSTD	
For the 7 difficulty items: Do you have difficulty with the following activities because of your vision? (yes, very great difficulties; yes, great difficulties; yes, some difficulties; no, no difficulties)						
1. Reading text in the newspaper	0.29 (0.1)	0.81	−9.9	0.76	−9.9	0.10
2. Recognizing faces of people you meet	0.95 (0.1)	1.29	9.9	1.28	9.9	−0.22
3. Seeing prices of goods when shopping	−0.14 (0.1)	0.89	−9.9	0.93	−6.0	−0.46
4. Seeing to walk on uneven ground	−0.09 (0.1)	1.25	9.9	1.39	9.9	−0.93
5. Seeing to do needlework and handicraft	−0.72 (0.1)	0.93	−6.1	0.89	−8.1	−0.25
6. Reading text on television	0.47 (0.1)	1.05	4.3	0.99	−0.7	0.17
7. Seeing to carry out a preferred hobby	0.31 (0.1)	0.99	−0.7	0.87	−7.8	0.13
Two global assessment items						
8. Do you experience that your present vision gives you difficulties in any way in your daily life? (yes, very great difficulties; yes, great difficulties; yes, some difficulties; no, no difficulties)	0.12 (0.1)	0.75	−9.9	0.70	−9.9	0.56
9. Are you satisfied or dissatisfied with your present vision? (very dissatisfied; rather dissatisfied; fairly satisfied; very satisfied)	−1.19 (0.1)	1.03	2.4	1.04	3.5	1.02

DIF = differential item functioning; MNSQ = mean square; ZSTD = standardized fit statistic

unidimensionality comes from PCA analysis of the residuals, which showed that variance explained by the measures was comparable for empirical calculation (64.2%) and by the model (64.2%). The unexplained variance explained by the first contrast was 1.6 Eigenvalue units (6.3%), which is close to the magnitude seen with random data. The 2 global assessment items correlated with the first contrast (satisfaction with vision 0.75 and overall visual difficulty 0.66), as did the reading item (0.20); however, the magnitude of the contrast was not enough to have much practical impact on the person measurement. The Cronbach's α was 0.91. The items were well targeted to the subjects (mean difference −1.21 logits preoperatively and postoperatively; −0.34 preoperatively and −2.32 postoperatively). This means that the difficulty of the items on the questionnaire were appropriate for the ability of patients. This is illustrated in the patient–item map shown in Figure 4. The 2 easiest questions were recognizing faces (0.95) and read text on television (0.47). The 2 most difficult questions were needlework and handicraft (−0.72) and satisfaction with vision (−1.19).

The preoperative Catquest-9SF scores were tested for stability over time. Linear regression showed a 0.05 logit reduction in score per year over the 11 years of registry data. This represents a small shift toward patients presenting for cataract surgery at a lower level of visual disability. This trend can be seen in Figure 5. The change in score was significantly different (ANOVA $F_{10862,10} = 11.52$; $P<.001$), with post hoc testing showing that the 1995 to 1998 scores

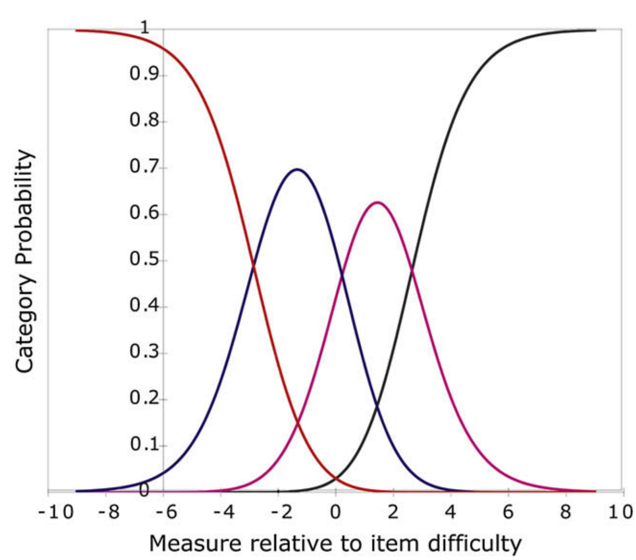

Figure 2. Category probability curves for the "satisfaction with vision" item.

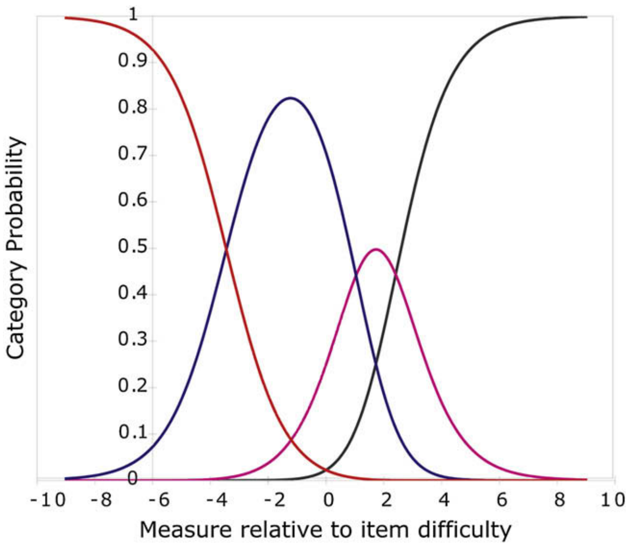

Figure 3. Category probability curves for the "difficulty in performing daily-life activities in general" item.

were significantly different from the 2001 to 2005 scores.

The Catquest-9SF was largely free of DIF. Two items showed a small level of DIF by sex; seeing to walk on uneven ground (0.64 logits) and seeing to do needlework and handicraft (0.62 logits) were both rated by women as easier relative to the other tasks. Four items showed a small level of DIF by age as follows: seeing to walk on uneven ground (rated 0.53 logits relatively easier in the 65- to 84-year group than in the <65-year group and rated 0.89 logits relatively easier in the >85-year group than in the <65 years group); seeing to do needlework and handicraft (0.56 logits relatively easier in the >85-year group than in the <65-year group); global assessment of difficulties (0.57 logits relatively more difficult in the 65- to 84-year group than in the <65-year group and 0.72 logits relatively more difficult in the >85-year group than in the <65-year group); and global assessment of satisfaction with vision (0.54 logits relatively more difficult in the >85-year group than in the <65-year group). There was neither DIF by the presence or absence of comorbidity nor by first-eye or second-eye cataract surgery status. For an instrument to be used on preoperative and postoperative populations, it is important that item functioning is consistent across measurement time frames. Therefore, DIF was tested between preoperative and postoperative data sets. Three items showed some DIF: seeing to walk on uneven ground (rated 0.93 logits relatively easier in the postoperative ranking); global assessment of difficulties (rated 0.56 logits relatively more difficult in the postoperative ranking); and global assessment of satisfaction with

```
           Patients - MAP - Items
      <more disabled>|<easier>
  5             .###  +
                      |
                      |
                  .#  |
                   .  |
  4                .  +
                   .  |
                  .# T|
                  .#  |
                   .  |
  3              .##  +
                 .##  |
                  .#  |
                 .##  |
                .###  |
  2             .###  +
              .#####  |
                  .# S|T
              .#####  |
              .#####  |
  1           .#####  +  recognizing faces
             .######  |S
              .#####  |
           .########  |  hobbies  walk on uneven ground  seeing text on TV
                  .#  |  reading newspaper
  0        .########  +M prices when shopping
              .#####  |  daily-life activities in general
              .##### M|
       .############  |  needlework and handicraft
                  .#  |S
 -1         .#######  +
              .#####  |
              .#####  |
              .#####  |T
              .#####  |  satisfaction with vision
 -2             .###  +
                .###  |
                .### S|
                .###  |
                  .#  |
 -3               .#  +
                .###  |
                  .#  |
                 .##  |
                   .  |
 -4                .  +
                  .# T|
                  .#  |
                   .  |
                   .  |
 -5                .  +
                  .#  |
                 .##  |
                      |
                      |
 -6              .##  +
      <less disabled>|<more difficult>
EACH '#' IS 61.
```

Figure 4. Person–item map of the 9-item short-form Catquest for the preoperative development group showing the distribution of Rasch-calibrated participant scores (*left*) and item locations (*right*). The items are well targeted to the patients as illustrated by the matching of the distributions (M = mean, S = 1 standard deviation, T = 2 standard deviations.

vision (rated 1.02 logits relatively more difficult in the postoperative ranking).

The Rasch analysis of the Catquest-9SF not only provided the item calibrations found in Table 2 but also item–category calibrations for each of the 4 response categories of the 9 items. These 36-item–category calibrations can be used as anchor values to convert ordinal category value to Rasch measurement estimates. This is valid for both preoperative and postoperative questionnaire data because the calibrations were developed using a combination of preoperative and postoperative data. Other investigators wishing to use the Catquest-9SF can use these calibrations to achieve Rasch measurement without the need to perform Rasch analysis. An Excel spreadsheet has been created for this purpose and is available from the authors.

Validation Phase

The preoperative and postoperative Catquest data from the second population (n = 10 886) were

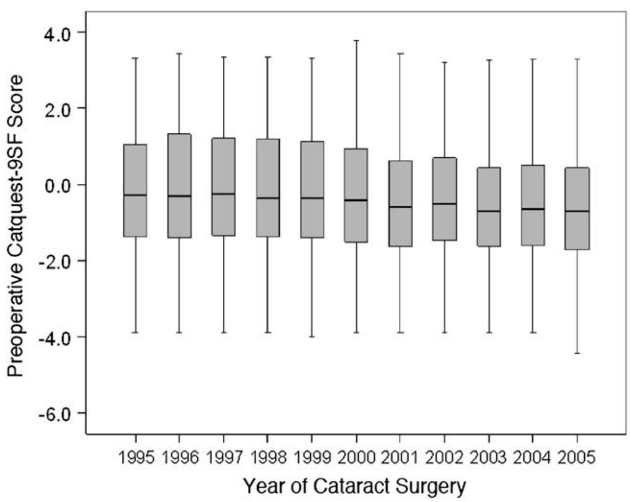

Figure 5. Box plot showing a slight trend toward lower levels of presenting visual disability over the decade of Swedish National Cataract Register data (boxes represent the middle half of the ranked data; black lines are the median values; bars indicate 95% confidence intervals).

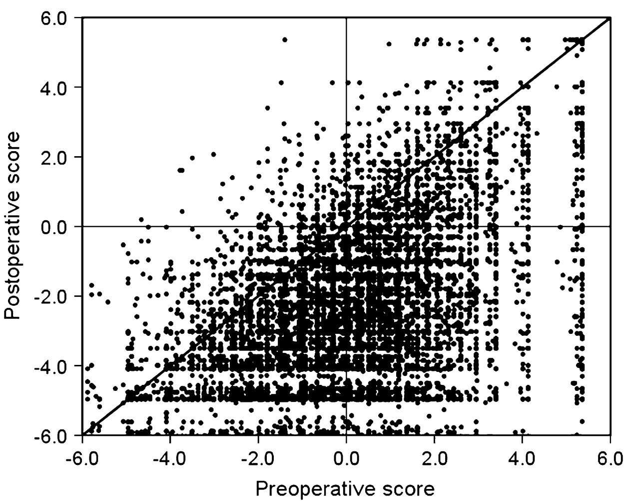

Figure 6. Scatterplot of preoperative Catquest-9SF score (logits) versus postoperative Catquest-9SF score (logits). The majority of cases improved, as shown by the points appearing below the 1:1 line.

converted to Rasch person estimated using the item–category anchor calibrations established in the development phase. These were correlated against preoperative and postoperative visual acuity data. The preoperative Catquest-9SF score correlated with visual acuity in the eye to be operated on ($r = 0.207$), visual acuity in the fellow eye ($r = 0.410$), and visual acuity in the better eye ($r = 0.431$). The postoperative Catquest-9SF score correlated with visual acuity in the operated eye ($r = 0.443$), visual acuity in the fellow eye ($r = 0.363$), and visual acuity in the better eye ($r = 0.476$).

The mean preoperative Catquest-9SF score was -0.32 ± 2.15 logits. The mean postoperative Catquest-9SF score was -3.21 ± 2.50 logits. This 3-logit improvement was statistically significant ($P < .0001$, paired 2-tailed t test). Figure 6 shows a scatterplot of the preoperative and postoperative Catquest-9SF scores. This shows that the majority of cases improved, as illustrated by the bulk of the data appearing below the 1:1 line; 9.8% of cases had a poorer Catquest-9SF score after cataract surgery than before surgery. The change in score with cataract surgery represents an effect size of 1.35.

DISCUSSION

The application of Rasch analysis to an existing conventionally developed questionnaire provides 2 distinct benefits. First, a greater insight into internal consistency is provided through the fit of the items to the model. In the case of Catquest, it was shown that the visual disability items were internally consistent and that the measurement of visual disability could be augmented through the addition of the 2 global assessment items because these were conceptually consistent with the same underlying trait. However, neither the symptoms items nor the frequency items were consistent with this trait, as illustrated by misfit to the model. This finding is consistent with research of quality-of-life instruments in which symptoms and disability failed to tap the same latent trait.[18,36] As described previously, the evaluation of original Catquest was made using a decision tree. The most important part was the 7 disability items. The Rasch analysis has confirmed that this part of the questionnaire represented the most valid measurement, which supports the role of these items in the original questionnaire. The refinement of the scale to purely measure visual disability is consistent with the World Health Organization International Classification of Functioning, Disability and Health definitions.[37]

The second benefit of Rasch analysis is the scoring of patient ability on a valid interval scale. This improves the precision of measurement by eliminating noise from nonlinearities in the summary scoring. Clearly, this results in more meaningful interpretation of scoring and also reduces the sample size required to find significant differences in clinical outcome studies.[14,15]

The finding that the frequency items did not contribute to the measurement of disability was not consistent with the theory behind the inclusion of these items in the original Catquest. It was expected that how often an activity was performed would give an indication of the importance of the activity and would therefore modulate visual disability. This expectation was not met by finding that including frequency of performing an activity in the same scale as visual disability actually adds noise to the measurement. Perhaps this indicates that people naturally take into account how often they perform an activity or how important an activity is to them when they assess how much difficulty they are having performing the activity.

The combination of perceived disabilities in daily-life and global assessment items formed the abridged version of Catquest with the greatest measurement precision. Therefore, we propose that the 9-item short form of the Catquest (Catquest-9SF) should be used in place of the original version. The Catquest-9SF has excellent precision and is unidimensional, as indicated by the fit statistics and PCA of the residuals. In addition to the questionnaire's internal consistency, the use of only 9 items greatly reduces the respondent burden, which makes the questionnaire suitable for use in clinical practice. Developing shorter questionnaires for broader implementation of patient-reported outcomes measurements is an often-pursued aim.[6,38,39] However, Rasch analysis has shown limitations in these questionnaires. It is difficult to have satisfactory measurement precision with a small number of items.[6] Attempts to create short-form questionnaires from existing questionnaires has arrived at minimum item sets of 12 items for the VDA, 16 items for the ADVS, and 10 items for the VF-14 (Pesudovs K, et al. IOVS 2005; 45:ARVO E-Abstract 3844).[5,7] However, both the ADVS and VF-14 showed poor targeting of item difficulty to patient ability, with the VF-10 requiring additional new items to optimize measurement. Not only can the Catquest-9SF measure with satisfactory precision, the items are well targeted to patient ability. Good targeting of a short-form visual disability instrument also occurs with the Cataract Outcomes Questionnaire (Pesudovs K, et al. IOVS 2005; 45:ARVO E-Abstract 3844). The better functioning of these short-form instruments compared with that of the ADVS and VF-14 is likely because the content of the long-form instruments is better targeted to patient ability, thus enabling a short form to be created while maintaining good targeting. Targeting is population dependent, and these data prove only that the Catquest-9SF is well targeted to the Swedish cataract population. Targeting should be tested for other populations, although the extremely large patient pool used in this study makes it unlikely that this is an atypical population for a developed country. The main issue for international adaptation is variation in the indications for cataract surgery. A reduction in the threshold visual impairment or disability for cataract surgery has been widely reported.[40] Although there has been a small shift toward patients presenting for cataract surgery at lower levels of visual disability over the 11 years of the data used in this study, the change is not large enough to adversely affect targeting of item difficulty to patient ability. Also, although optimal targeting occurs for the preoperative population, in postoperative cases, the items are too easy for the patients. This is an acceptable compromise in a cataract-related visual disability instrument because the goal of cataract surgery is to eliminate visual disability; therefore, by definition, targeting will change. This can lead to a ceiling effect. Figure 6 shows that some postoperative score are at the ceiling; thus, there must be some measurement distortion for these individuals. However, many postoperative patients are well measured without distortion. That the Catquest-9SF can still discriminate postoperative patients with good precision is an important indicator of its validity for clinical measurement of cataract outcomes.

Although the large sample size in this study has several advantages, it placed several limitations on the analyses. Significance-based testing was not possible because the extreme sample size provided so much power that any level of misfit would be identified as significant.[30] Nevertheless, by using magnitude as the indicator of error, the Catquest-9SF was shown to be largely free of DIF and to be unidimensional.

The Rasch scaling of the Catquest-9SF provides for a spreadsheet-based conversion of raw data to Rasch estimates. Therefore, clinicians who wish to measure cataract surgery outcomes using Catquest-9SF do not have to apply Rasch analysis to obtain the benefits of true interval data, such as suitability for parametric statistical analysis. The development of this conversion algorithm in such a large population suggests that the model will be very stable. This method is robust to even large amounts of missing data.

The Catquest-9SF was shown to correlate well with visual acuity in the better eye. This is consistent with findings in other studies that show that better-eye visual acuity correlates best with visual disability measurement.[41] Moreover, the correlations with visual acuity here are higher than for most questionnaires, which is likely due to the high precision of the Catquest-9SF.[42,43] The Catquest-9SF was highly responsive to cataract surgery. The effect size reported here can be considered to be large; convention holds that effect sizes of 0.20 to 0.49 are considered small, 0.50 to 0.79 are moderate, and 0.80 or above are large.[44]

The good characteristics of Catquest-9SF have led to its incorporation in the Swedish National Cataract Register. Future work using the Catquest-9SF will include a more comprehensive evaluation of the outcomes of cataract surgery. The relative benefit of first-eye and second-eye cataract surgery and the role of cofactors such as ocular comorbidity, age, sex, and location will be reported in a subsequent manuscript.

The importance of Rasch analysis in the development and scoring of questionnaires has been recognized in standards proposed for the assessment of questionnaire quality.[16,45,46] Numerous calls have been made in ophthalmology for the development of instruments using this technology.[11,47,48] Clearly, a Rasch-scaled questionnaire should be used wherever possible for the measurement of outcomes.[48,49] The Catquest-9SF, along with the Cataract Outcomes Questionnaire, represent the state of the art for the measurement of visual disability, and is likely superior to the Rasch-analyzed versions of the ADVS and the VF-14 and certainly superior to any visual disability instruments not subjected to Rasch analysis (Pesudovs K, et al. IOVS 2005; 45:ARVO E-Abstract 3844).

In conclusion, the Catquest-9SF is a valid short-form visual disability instrument that is ideal for the measurement of patient-reported outcomes of cataract surgery. From a clinical viewpoint, the advantages with this instrument include that it measures disability, gives interval scale scoring, has high precision, is short and minimizes response burden, is sensitive to changes after cataract surgery, has high effect size, and has good targeting.

REFERENCES

1. 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; 30:1111–1126
2. Steinberg EP, Tielsch JM, Schein OD, Javitt JC, Sharkey P, Cassard SD, Legro MW, Diener-West M, Bass EB, Damiano AM, Steinwachs DM, Sommer A. The VF-14; an index of functional impairment in patients with cataract. Arch Ophthalmol 1994; 112:630–638
3. Mangione CM, Orav EJ, Lawrence MG, Phillips RS, Seddon JM, Goldman L. Prediction of visual function after cataract surgery; a prospectively validated model. Arch Ophthalmol 1995; 113:1305–1311
4. Schein OD, Steinberg EP, Cassard SD, Tielsch JM, Javitt JC, Sommer A. Predictors of outcome in patients who underwent cataract surgery. Ophthalmology 1995; 102:817–823
5. Velozo CA, Lai JS, Mallinson T, Hauselman E. Maintaining instrument quality while reducing items: application of Rasch analysis to a self-report of visual function. J Outcome Meas 2000–2001; 4:667–680
6. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient α in the creation of minimum item sets. Med Care 2004; 42(suppl 1):I17–I24
7. 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. Available at: http://www.iovs.org/cgi/reprint/44/7/2892. Accessed December 9, 2008
8. Lundström M, Roos P, Jensen S, Fregell G. Catquest questionnaire for use in cataract surgery care: description, validity, and reliability. J Cataract Refract Surg 1997; 23:1226–1236
9. Lundström M, Stenevi U, Thorburn W. The Swedish National Cataract Register: a 9-year review. Acta Ophthalmol Scand 2002; 80:248–257
10. Lundström M, Stenevi U, Thorburn W, Roos P. Catquest questionnaire for use in cataract surgery care: assessment of surgical outcomes. J Cataract Refract Surg 1998; 24:968–974
11. Pesudovs K. Patient-centred measurement in ophthalmology—a paradigm shift. BMC Ophthalmol 2006; 6:25. Available at: http://www.biomedcentral.com/content/pdf/1471-2415-6-25.pdf. Accessed December 9, 2008
12. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. Arch Phys Med Rehabil 1989; 70:857–860
13. Fisher WP Jr, Eubanks R, Marier RL. Equating the MOS SF36 and the Lsu HSI physical functioning scales. J Outcome Meas 1997; 1:329–362
14. 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(suppl 1):I25–I36
15. 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
16. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. Optom Vis Sci 2007; 84:663–674
17. Lamoureux EL, Pallant JF, Pesudovs K, Hassell JB, Keeffe JE. The Impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. Invest Ophthalmol Vis Sci 2006; 47:4732–4741. Available at: http://www.iovs.org/cgi/reprint/47/11/4732. Accessed December 9, 2008
18. 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
19. 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. Available at: http://www.iovs.org/cgi/reprint/47/7/2789. Accessed December 9, 2008
20. Lundström M, Stenevi U, Thorburn W. Cataract surgery in the very elderly. J Cataract Refract Surg 2000; 26:408–414; erratum, 635
21. Lundström M, Stenevi U, Thorburn W. Quality of life after first- and second-eye cataract surgery; five-year data collected by the Swedish National Cataract Register. J Cataract Refract Surg 2001; 27:1553–1559
22. Lundström M, Brege KG, Florén I, Lundh B, Stenevi U, Thorburn W. Cataract surgery and quality of life in patients with age related macular degeneration. Br J Ophthalmol 2002; 86:1330–1335
23. Lundström M, Albrecht S, Nilsson M, Åström B. Benefit to patients of bilateral same-day cataract extraction: randomized clinical study. J Cataract Refract Surg 2006; 32:826–830
24. Rasch BG. Probabilistic Models for Some Intelligence and Attaiment Tests. Copenhagen, Denmark, Danmarks Paedogogiske Institut, 1960
25. Linacre JM. A User's Guide to Winsteps: Rasch-Model Computer Program. Chicago, IL, Mesa Press, 2002
26. Linacre JM. Winsteps Rasch measurement computer program. Chicago, IL, Winsteps, 2006. Available at: http://winsteps.com. Accessed December 9, 2008
27. Andrich D. A rating scale formulation for ordered response categories. Psychometrika 1978; 43:561–573
28. Wolfe EW, Chiu CW. Measuring pretest–posttest change with a Rasch rating scale model. J Outcome Meas 1999; 3:134–161
29. Smith RM. Person fit in the Rasch model. Educ Psychol Measurement 1986; 46:359–372
30. Linacre JM. Size vs. significance: infit and outfit mean-square and standardized chi-square fit statistic. Rasch Measurement

Trans 2003; 17:918. Available at: http://www.rasch.org/rmt/rmt171n.htm. Accessed December 8, 2008
31. Tennant A, Pallant JF. Unidimensionality matters! (A tale of two Smiths?) Rasch Measurement Trans 2006; 20:1048–1051. Available at: http://www.rasch.org/rmt/rmt201c.htm. Accessed December 8, 2008
32. Smith RM, Miao CY. Assessing unidimensionality for Rasch measurement. In: Wilson M, ed, Objective Measurement: Theory into Practice, vol 2. Norwood, NJ, Ablex, 1994; 316–328
33. Wright BD, Douglas GA. Best test design and self-tailored testing. MESA Research Memorandum No. 19. Statistical Laboratory, Department of Education. Chicago, IL, University of Chicago, 1975
34. Wright BD, Douglas GA. Rasch item analysis by hand. MESA Research Memorandum Number 21. Statistical Laboratory, Department of Education. Chicago, IL, University of Chicago, 1976
35. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. Health Qual Life Outcomes April 11, 2003; 1:4. Available at: http://www.hqlo.com/content/pdf/1477-7525-1-4.pdf. Accessed December 9, 2008
36. Massof RW, Fletcher DC. Evaluation of the NEI visual functioning questionnaire as an interval measure of visual ability in low vision. Vision Res 2001; 41:397–413
37. WHO International Classification of Functioning, Disability and Health. Geneva, Switzerland, World Health Organization, 2001 (ICIDH-2)
38. Pesudovs K, Elliott DB. Shortening the VF-14 visual disability questionnaire [letter]. J Cataract Refract Surg 2006; 32:6
39. Uusitalo RJ, Brans T, Pessi T, Tarkkanen A. Evaluating cataract surgery gains by assessing patients' quality of life using the VF-7. J Cataract Refract Surg 1999; 25:989–994
40. Leinonen J, Laatikainen L. Changes in visual acuity of patients undergoing cataract surgery during the last two decades. Acta Ophthalmol Scand 2002; 80:506–511
41. Mitchell J, Wolffsohn J, Woodcock A, Anderson SJ, Ffytche T, Rubinstein M, Amoaku W, Bradley C. The MacDQoL individualized measure of the impact of macular degeneration on quality of life: reliability and responsiveness. Am J Ophthalmol 2008; 146:447–454
42. Acosta-Rojas ER, Comas M, Sala M, Castells X. Association between visual impairment and patient-reported visual disability at different stages of cataract surgery. Ophthalmic Epidemiol 2006; 13:299–307
43. Alonso J, Espallargues M, Andersen TF, Cassard SD, Dunn E, Bernth-Petersen P, Norregaard JC, Black C, Steinberg EP, Anderson GF. International applicability of the VF-14; an index of visual function in patients with cataracts. Ophthalmology 1997; 104:799–807
44. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000; 53:459–468
45. de Boer MR, Moll AC, de Vet HCW, Terwee CB, Völker-Dieben HJM, Van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: a systematic review. Ophthalmic Physiol Opt 2004; 24:257–273
46. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60:34–42
47. Spaeth G, Walt J, Keener J. Evaluation of quality of life for patients with glaucoma. Am J Ophthalmol 2006; 141(suppl):S3–S14
48. Weisinger HS. Assessing the value of LASIK by patient-reported outcomes using quality of life assessment [letter]. J Refract Surg 2006; 22:14–15; reply by H-Y Kang, 15
49. Lundström M, Wendel E. Assessment of vision-related quality of life measures in ophthalmic conditions. Expert Rev Pharmacoeconomics Outcomes Res 2006; 6:691–724

First author:
Mats Lundström, MD, PhD

EyeNet Sweden, Blekinge Hospital, Karlskrona, Sweden