Using Rasch analysis to revisit the validity of the Cataract TyPE Spec instrument for measuring cataract surgery outcomes

Vijaya K. Gothwal, PhD, Thomas A. Wright, Ecosse L. Lamoureux, PhD, Konrad Pesudovs, PhD

PURPOSE: To assess the psychometric properties of the Cataract TyPE Specification (Spec) questionnaire using the Rasch model.

SETTING: Flinders Medical Centre, Adelaide, Australia.

METHODS: The 12-item Cataract TyPE Spec questionnaire was self-administered to patients drawn from the cataract surgery waiting list. The questionnaire and its 5 subscales were assessed for fit to the Rasch model. Response category performance, item-fit targeting, unidimensionality (using principal components analysis), and differential item functioning were assessed. A shortened version (11 items) was tested for criterion validity by determining correlation with a global rating of vision question.

RESULTS: Two hundred ninety-four patients responded to the questionnaire. The response categories for each question functioned as intended. Person-separation reliability was high (0.90). Deletion of 1 misfitting item (nighttime driving) improved overall model fit. The principal components analysis of the residuals demonstrated unidimensionality for the 11-item Cataract TyPE Spec and 2 subscales. However, items were targeted to a less able population. Only 2 subscales (near vision and glare) were valid. There was a good statistically significant correlation between the Likert-scored global rating of vision and the Rasch-scaled Cataract TyPE Spec score ($r = -0.66$, $P<.0001$), suggesting good criterion validity.

CONCLUSIONS: With minor modifications, the Cataract TyPE Spec questionnaire and its glare and near-vision subscales were good measures of visual functioning in the cataract patient. Additional items to suit the more able, including patients having second-eye surgery, could improve the measurement properties.

J Cataract Refract Surg 2009; 35:1509–1517 © 2009 ASCRS and ESCRS

Cataract is the leading cause of treatable blindness in the world, and cataract surgery is expected to increase with the aging population.[1] It is the most commonly performed procedure in Australia.[2,3]

The need to include patient-reported outcomes (or questionnaires) in the assessment of cataract surgery is now widely appreciated.[4–6] Several questionnaires developed for this purpose, such as the Activities of Daily Vision Scale (ADVS),[7,8] Visual Disability Assessment (VDA),[9] Visual Function 14 (VF-14),[10] Catquest,[11] and Cataract TyPE Specification (TyPE Spec),[12,13] have used the traditional psychometric approach; that is the classical test theory. However there are 2 important limitations of classical test theory: the validity of the scoring method and the validity of the items included. The scoring problem of classical test theory arises from the assumption that distances on the scales are equal over the full range of the scale and the scale is erroneously treated as an interval scale based on the ordinal level of scoring.[12,13] More recently, modern psychometric approaches have been adopted such as the Rasch measurement models.[14–16] Rasch analysis provides unparalleled insight into content validity and targeting of item difficulty to patient ability, not possible with classical test theory.[17,18] Although none of the aforementioned questionnaires (ADVS, VDA, VF-14, Catquest) have been developed using Rasch Analysis, they have been tested and revalidated using Rasch analysis (Pesudovs K, et al. IOVS 2005; 46:ARVO E-Abstract 3844. Available at: http://abstracts.iovs.org/cgi/content/abstract/46/5/3844. Accessed May 24, 2009).[19–21]

The Cataract TyPE Spec is a reliable and valid scale, by classical test theory, for measuring vision-related

Published by Elsevier Inc.
0886-3350/09/$—see front matter
doi:10.1016/j.jcrs.2009.03.056

function and outcomes of cataract extraction.[12,13] In view of the limitations of the classic test theory, we sought to use Rasch analysis to revalidate the Cataract TyPE Spec instrument in an Australian cataract population waiting to have cataract extraction. This included, if necessary, making revisions to the content of the instrument to improve its validity. However, the main aim was to use Rasch analysis to estimate linear interval measures from ordinal raw scores. This improves the accuracy of scoring and removes measurement noise,[22–24] thus improving sensitivity to intervention-related changes.[22–25] Linear scoring also enables parametric statistics to be validly applied to questionnaire scores. The final aim was to make Rasch scoring available to all investigators using the Cataract TyPE Spec questionnaire through a simple spreadsheet-based scoring conversion.

PATIENTS AND METHODS

Questionnaire

The Cataract TyPE Spec questionnaire assesses visual functioning in 5 dimensions: distance vision, near vision, daytime driving, nighttime driving, and glare (Table 1). There are 5 response options for the perceived difficulty levels for both vision-related and glare-related questions (summary scoring): totally disabled 5, quite a lot 4, some difficulty 3, a little bit of difficulty 2, and not at all 1. An additional option is provided if the subject does not do an activity for other reasons, which is scored as zero, and is considered as missing data for Rasch analysis. One question relates to the global rating of vision. This question was used to assess criterion validity.

Submitted: January 27, 2009.
Final revision submitted: none.
Accepted: March 30, 2009.

From the NH&MRC Centre for Clinical Eye Research (Gothwal, Wright, Pesudovs), Department of Ophthalmology, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, South Australia, the Centre for Eye Research Australia (Lamoureux), Department of Ophthalmology, University of Melbourne, Victoria, and Vision CRC (Lamoureux), Sydney, Australia; Meera and LB Deshpande Centre for Sight Enhancement (Gothwal), Vision Rehabilitation Centres, L.V. Prasad Eye Institute, Hyderabad, India; Singapore Eye Research Institute (Lamoureux), Singapore National Eye Centre, Singapore.

No author has a financial or proprietary interest in any material or method mentioned.

Supported in part by National Health and Medical Research Council, Centre of Clinical Research Excellence Grant 264620, Canberra, Australia.

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

There are also questions about other characteristics, such as recent illness, injury, or emotional upset, and the help required (if any) to fill out the questionnaire. These latter questions were meant to be used as demographic variables, not to evaluate the benefits of cataract surgery. Therefore, these questions were not included in the analysis.

Patients

Patients with cataract were drawn from the surgical waiting list (average waiting period 3 to 4 months) of the Ophthalmology Department, Flinders Medical Centre, Adelaide, South Australia. All patients had previously had a comprehensive ocular examination and were diagnosed to have cataract as the principal cause of visual disability and who required surgical intervention. Consecutive patients on the list were invited to participate. This included patients awaiting first- or second-eye surgery and those with ocular comorbidity (eg, glaucoma, diabetic retinopathy). It is important to include these comorbidities as it is unlikely to find a sample of an elderly cataract population without coexisting ocular conditions. Other inclusion criteria were age 18 years or older, no severe cognitive impairment, and ability to converse in English without the need for an interpreter. Ethical approval was obtained, and all patients who agreed to participate signed a consent form. This research adhered to the tenets of the Declaration of Helsinki.

Clinical Assessment

Visual acuity assessments were performed for each eye separately and binocularly using a computerized presentation of letters following logMAR principals with a screen luminance of 150 candelas/m^2 and letter-by-letter scoring. Only binocular visual acuity measurements were used for analysis because these are more closely related to disability than better-eye or worse-eye measurements.[26,27]

Rasch Analysis

The Cataract TyPE Spec data were fitted to the Rasch rating scale model[28] using the Winsteps Rasch measurement software[29,30] (version 3.66, Winsteps). The item response theory, specifically Rasch models, has been described in detail.[31] Rasch models are probabilistic mathematical models that have been used in the validation of a series of vision-specific questionnaires.[19,32–34] In brief, Rasch models attempt to estimate the values of latent variables on an interval scale from item scores that form an ordinal scale. In the Rasch model of visual disability, disability can be considered to lie on a ruler, similar to an ordinary ruler, where "no disability" is anchored at one end and "maximum disability" at the other end. The range of disability is expressed in log-odds probability units (logits). The longer the scale length, the better it is at representing the visual disability. An item (question) difficulty represents the position in logits that the item occupies on the linear disability scale; the center of the scale is set at zero (ie, mean item difficulty). Similarly, person (participant) ability represents the location of the person on the same scale in logits. For the Cataract TyPE Spec instrument, a positive logit score indicates a less difficult item and less able participant (ie, the participant has lower ability than that required to endorse the item).

First, the participant's use of the rating scale was examined using the category probability curves. For a well-fitting item, one would expect that across the whole range of the

Table 1. Items included in the original version of the Cataract TyPE Spec questionnaire.

Item	Description	Subscale
2	Vision hinders usual activity	Distance vision
3	Vision hinders recognising people	Distance vision
4	Vision hinders reading price labels	Near vision
5	Vision hinders reading magazine	Near vision
6	Vision hinders knitting	Near vision
7	Vision hinders watching television	Distance vision
8	Vision hinders daytime driving	Daytime driving
9	Vision hinders nighttime driving	Nighttime driving, glare
10	Glare hinders usual activities	Glare
11	Glare hinders reading shiny paper	Glare, near vision
12	Glare hinders driving toward the sun or oncoming headlights	Glare, daytime driving, nighttime driving
13	Glare hinders walking outside on a sunny day	Glare
—	Global rating question*	—

*The original version included a global rating question: "How do you rate your vision?" This was not included in the score.

trait being measured (visual disability in this case), the higher response option (eg, quite a lot or totally disabled) would show higher probability of endorsement than the lower response options (eg, a little bit of difficulty or not at all). One of the most common source of items that do not fit is the participant's inconsistent use of these response options, which causes disordered thresholds. The threshold represents the equal probability point between any 2 adjacent categories. If disordering occurs, combining adjacent categories can often, although not always, solve the problem and improve overall model fit. However, the newly formed categories remain theoretical and must be tested before being recommended.[35,36]

Second, the overall reliability of the Cataract TyPE Spec instrument was estimated by examining the person-separation statistics and person-separation reliability. Person separation is a measure of spread in the test sample, and reliability is a measure of the true variance to the observed measure variance. In general, reliability ranges from 0.0 to 1.0, with values higher than 0.8 (or person separation >2.0) considered the minimum acceptable level that indicates the ability of the instrument to distinguish 3 strata of participants (high, average, low) and that represents a useful range of task difficulty.[28,37] Model fit must be viewed in conjunction with response scale functioning and item fit because problems in these 2 measurement properties can decrease overall model fit.

Third, compliance of the fit of the observed data to the model expectations was assessed. Because the Rasch model is probabilistic and not deterministic, some failure of the model to predict the observed values is expected. Two statistics are used to represent these deviations: infit mean square (information-weighted fit statistic) and outfit mean square (outlier-sensitive fit statistic). Items that fit perfectly to the unidimensional scale have an expected infit and outfit mean square of 1.0. Although acceptable levels of these fit statistics vary,[38] values of 0.7 to 1.3 (30% less or more variance than expected) were considered acceptable in the present study. Items with poorer fit statistics (ie, misfitting items) usually reflect a problem with coherence of the item with the underlying trait, and they are considered "noisy."[39] Testing for item fit forms 1 of the assessments of unidimensionality. If items do not fit the model's expectations, unidimensionality is not preserved.

Further examination of unidimensionality (a primary requirement for measurement) was assessed using a principal component analysis of the residuals.[30,40] Unidimensionality refers to the capacity of the instrument to measure the specific attributes or underlying trait. Two criteria to were used to confirm unidimensionality. First, the proportion of variance explained by the Rasch measure should be comparable for empirical calculation as well as that explained by the model.[41] The second criterion is an eigenvalue less than 2.0 of the unexplained variance explained by the first contrast.[42]

Differential item functioning (DIF) concerns the expectation that participants who are in different groups but have equal levels of visual disability would have the same probability of selecting a particular response.[43,44] Differential item functioning analyses were used to identify items potentially biased by age (<75 years versus ≥75 years), sex, systemic and ocular comorbidity, and first versus second eye surgery. The definition of DIF was based on magnitude as follows: insignificant DIF, <0.50 logits; mild (but probably inconsequential), between 0.50 and 1.00 logits; and notable, >1.00 logits.[45,46] These DIF variables were selected a priori in the present study. Age was included because cataract surgery is performed over a wide age range. Sex was included because some of the items may be easier to perform for women than for men. Participants with comorbidities (systemic or ocular) may find some items more difficult than the rest. Participants who were awaiting cataract surgery in the second eye may find some items easier to perform than those awaiting surgery in the first eye.

In addition to the measurement properties described above, the Winsteps program enables item difficulty and person ability to be visualized along a linear scale (like a ruler), which is known as a person–item map. Such a map can be used 3 ways; that is, to determine (1) the extent to which item positions match person positions (targeting) (if positions do not line up, items are likely inappropriate [eg, too easy or too hard] for the persons); (2) whether there are gaps in the measure, which if present indicate the need for more items; (3) an item hierarchy, which provides information about the most and least difficult items and more and less able persons. For a valid measurement, a well-targeted scale must have adequate spread along the dimensions of measurement with negligible floor and ceiling effects.[47]

The proposed subscales of the Cataract TyPE Spec instrument were analyzed separately using a rigorous approach similar that used for the entire instrument because subscales have the same requirements as whole scales for sufficiency of psychometric properties.

Criterion validity was assessed by determining the relationship between the Likert score from the question on global rating of vision and the Rasch-scaled score for the overall Rasch-scaled Cataract TyPE Spec score as well as for the subscales. The relationship between visual acuity and the Rasch-scaled score was also assessed.

Summarystatistics and correlations for criterion validity were generated with SPSS statistical analysis software (version 16, SPSS, Inc.). A P value less than 0.05 was considered statistically significant.

RESULTS

Two hundred ninety-four patients responded to the visual disability scale. The 12 items were initially fitted to the Rasch model. Table 2 shows the patients' characteristics.

Assessment of the Rating Scale and Reliability

Figure 1 shows the category probability curves for an item of the Cataract TyPE Spec questionnaire. There was no evidence of disordered thresholds with the 5-category response scale; therefore, the originally proposed response scale was retained.

Table 2. Participant characteristics (N = 294).

Characteristic	Result
Mean age (y) ±SD	74.6 ± 9.5
Sex, n (%)	
Male	123 (41.8)
Female	171 (58.2)
Binocular visual acuity	
Mean ± SD	
LogMAR	0.27 ± 0.20
Snellen	$6/12^{+}$
Range	
LogMAR	-0.14 to 1.00
Snellen	$6/4.8^{+2}$ to 6/60
Awaiting second-eye surgery, n (%)	118 (41.9)
Ocular comorbidity,* n (%)	
Present	136 (47.9)
Absent	148 (52.1)
Duration of cataract (y)	
Median	1
Range	0 to 80
Systemic comorbidity,† n (%)	
Present	258 (93.1)
Absent	19 (6.9)

*Includes, for example, glaucoma, diabetic retinopathy, and age-related macular degeneration as well as data missing for 10 cases

†Includes, for example, diabetes, hypertension, and angina as well as data missing for 17 cases

The overall fit of the data to the model was good, indicating that overall, the 12-item scale formed a valid measure. The real person separation was 2.95 and the reliability, 0.90.

Item Fit

One item ("How much does your vision hinder, limit, or disable you in nighttime driving?") showed significant misfit (infit mean square 1.36). After the item was deleted, the infit mean square was within the acceptable range for the remaining 11 items, with little loss of person separation (2.88).

Targeting

Figure 2 shows the person–item map for the revised 11-item Cataract TyPE Spec questionnaire. The mean person ability was −1.47 logits ± 1.95 (SD). The range of person ability (5.34 to −6.22 logits) was not significantly different from a normal distribution (Kolmogorov-Smirnov Z test score 0.88, $P = .41$). There was an uneven spread of items along the visual disability continuum. On the whole, the items were too easy for the abilities of the patients. Further inadequacies in the measurement were reflected by redundancy of the items; that is, pairs of items (eg, "reading price labels in shops and supermarkets" and "reading shiny paper") were located at the same difficulty level, indicating that they were addressing a similar activity. Redundancy is acceptable only if the aim of the instrument is high precision rather than brevity. The

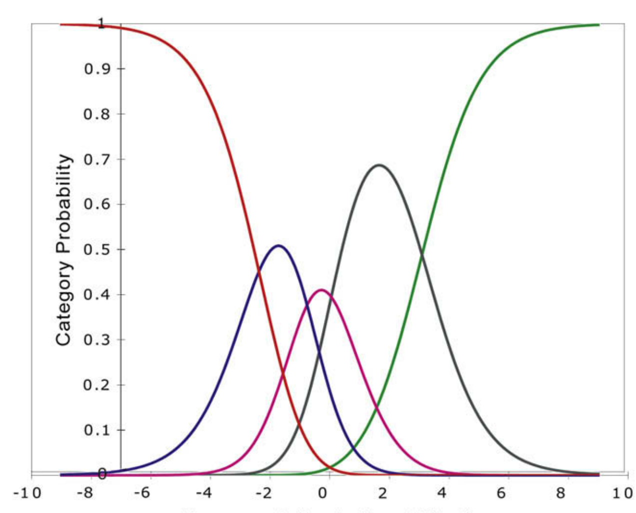

Figure 1. Category probability curves for the 11-item Cataract TyPE Spec questionnaire showing the range over which each of the 5 categories is most likely to be chosen. Boundaries occur at points along the scale where the category most likely to be chosen changes from one to the next.

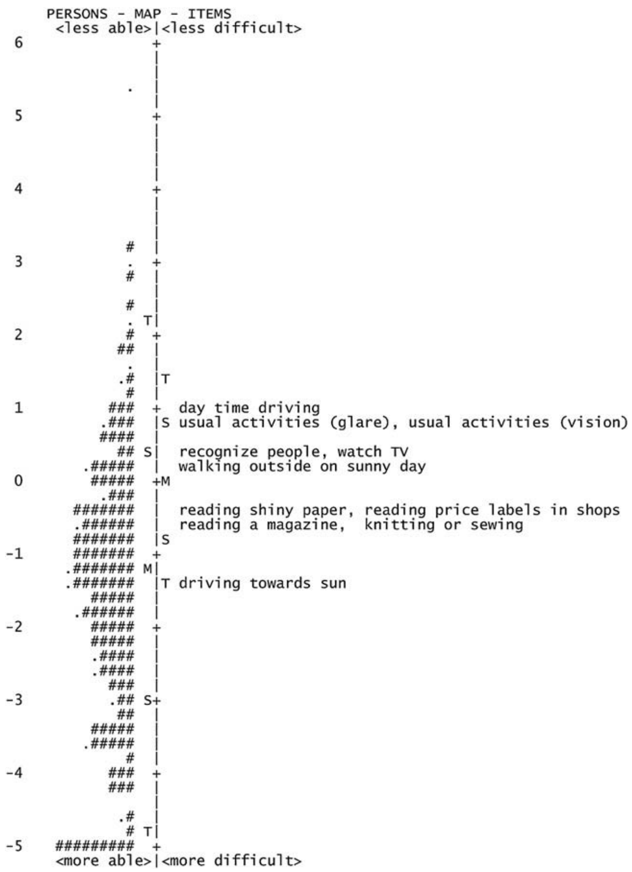

Figure 2. Person–item map for 11-item Cataract TyPE Spec questionnaire. The subjects are on the left of the dashed line, and more able subjects are located at the bottom of the map. Items are located on the right of the dashed line and more difficult items are located at the bottom of the map. Each "#" represents 2 subjects (M = mean; S = 1 standard deviation from the mean; T = 2 standard deviations from the mean).

addition of items representing more difficult activities would improve targeting of the instrument.

The 3 most difficult items were "driving toward the sun or oncoming headlights," "knitting or sewing," and "reading a magazine, newspaper, or books." Conversely, the 3 least difficult items were "daytime driving," "your usual daily activities (vision related)," and "your usual daily activities (glare related)" (Table 3).

Unidimensionality

The principal components analysis of the residuals indicated that the variance explained by the measures was comparable for the empirical calculation (64.2%) and by the model (65.0%). The unexplained variance explained by the first contrast was 2.0 eigenvalue units, which is close to the magnitude seen with random data. Taken together, these findings confirmed the unidimensionality of the Cataract TyPE Spec instrument.

Differential Item Functioning

The Cataract TyPE Spec questionnaire was largely free of DIF. However, 3 items showed a minimal level of DIF by sex, with male participants rating 1 item as easier relative to other items: "vision hinders, limits, or disables you in your usual activities" (0.58 logits). Female participants, on the other hand, rated the following 2 items as easier than the male participants: "knitting or sewing" (0.54 logits) and "walking outside on a sunny day" (0.71 logits). Participants who did not have ocular comorbidity rated the item "driving toward the sun" as 0.53 logits easier than those who had ocular comorbidity.

Performance of Subscales Within the Rasch Model

Table 4 shows the results of subscale analyses. Of the 5 proposed subscales, only glare and near vision fit the Rasch model. Given the low level of discrimination of the remaining subscales, they did not form valid scales. Furthermore, no amount of modifications could improve the overall functioning of these subscales to a satisfactory level.

Glare

The person separation and reliability for glare fell within the recommend range (Table 4). One item (nighttime driving) misfit (infit mean square 1.45). After its deletion, the remaining 4 items fit and the person separation (2.11) remained largely unaffected. However, the targeting worsened (−1.34 logits). The principal components analysis of the residuals showed unidimensionality. There was minimal DIF by sex for 3 items. Male participants rated the following 2 items as relatively easier compared with female participants: "glare hinders, limits, or disables your usual daily activities" (0.61 logits) and "read shiny paper" (0.53 logits). Females, on the other hand, rated the item "walking outside on a sunny day" 0.87 logits easier compared with the male participants. This subscale formed a valid measure. However, similar to the full version of the Cataract TyPE Spec instrument, this subscale was also targeted toward the less able end of the population, with more able subjects having little or no difficulty with more difficult items.

Near Vision

The person separation and reliability for near vision were within the acceptable range (Table 4). There were no misfitting items. In contrast to the glare subscale, the items in this subscale were better targeted to the participants' abilities. The principal components

Table 3. Item-fit statistics from the 11-item Cataract TyPE Spec questionnaire fitted to Rasch model.

TyPE Item	Description	Item Calibration	Standard Error	Infit Mean Square	Outfit Mean Square
8	Day time driving	1.04	0.13	1.17	0.88
2	Usual daily activities (vision related)	0.88	0.09	0.92	0.87
10	Usual daily activities (glare related)	0.71	0.09	0.96	0.90
7	Watching television	0.49	0.09	1.01	1.07
3	Recognize people	0.40	0.09	1.27	1.22
13	Walking on sunny day	0.15	0.09	1.06	1.05
4	Reading price labels	−0.45	0.09	0.87	0.82
11	Reading shiny paper	−0.47	0.08	0.91	0.94
5	Reading a magazine	−0.66	0.08	0.78	0.75
6	Knitting or sewing	−0.69	0.10	1.17	1.14
12	Driving toward sun	−1.41	0.10	1.17	1.16

analysis of the residuals confirmed unidimensionality. No items showed DIF. This subscale also formed a valid measure. Similar to the full version of the Cataract TyPE Spec instrument and the glare subscale, this subscale was targeted toward the less able end of the population, with more able subjects having little or no difficulty with more difficult items.

Criterion Validity

The Spearman rank correlation of visual acuity with the 11-item Rasch scaled Cataract TyPE Spec instrument was 0.25. There was a good statistically significant correlation ($r = -0.66$, $P < .0001$) between the global rating of vision and the 11-item Rasch scaled Cataract TyPE Spec instrument. The Spearman rank correlation between the global rating of vision and the glare subscale was −0.41 and between the global rating of vision and the near-vision subscale, −0.57 (both $P < .0001$).

DISCUSSION

The Cataract TyPE Spec questionnaire performed well in our population. It was highly discriminating, as evidenced by the person separation for the 12-item Cataract TyPE Spec and the revised 11-item version, indicating that the instrument can discriminate between 4 strata of participant ability. Our findings of high precision are in line with those reported for Catquest-9SF,[21] another Rasch-revalidated questionnaire for a cataract population. There was no need to collapse categories, supporting the originally proposed 5-category response scale for the Cataract TyPE Spec questionnaire. Our finding of ordered thresholds for a 5-category response scale in a cataract population is consistent with the 10-item Vision Core Measure 1 scale.[48] Although, other visual functioning questionnaires used in cataract patients required fewer response categories,[19] this difference probably reflects the appropriateness of category labels for the Cataract TyPE Spec questionnaire. One reason for the optimum use of the categories in this study could be because the categories ("not at all," "a little bit") are biased toward more able patients (ie, the less affected end of scale). However, despite ordered thresholds, there was underuse of the extreme response category (ie, totally disabled, 4%); this reflects the finding of suboptimum targeting in our study population.

There was a significant floor effect in that the items did not target participants with low levels of visual disability (ie, at bottom of the map). Although there were participants with low levels of visual disability (logits > −1.0), no items had difficulty estimates in that area. However, this may not entirely be the case

Table 4. Results of testing of fit of TyPE Spec subscales to the Rasch model.

	Subscale				
Parameter	Glare	Near Vision	Daytime Driving	Nighttime Driving	Distance Vision
Items (n)	5	4	2	2	3
Misfitting items (n)	1	0	0	0	0
Person separation	2.10	2.02	1.21	1.40	1.41
Person separation reliability	0.81	0.80	0.59	0.66	0.67
Mean item location	0	0	0	0	0
Mean person location	−0.87	−0.88	−2.79	−0.56	−1.69
Principal components analysis (eigenvalue)	1.6	1.5	—	—	1.7

because items use multiple response categories and each item is represented not just by its mean value by but levels of endorsement as well. Such intricate details remain obscure in the person–item map. Nevertheless, the restricted range of item difficulty (thresholds) indicates the Cataract TyPE Spec scale's representation of the visual disability construct is skewed, suggesting less than optimum targeting of the cataract patients to the questionnaire's items and thresholds. The participants were mistargeted by 1.47 logits. A little more than half the patients (59.9%) were awaiting surgery in their first eye, and our findings of poor targeting are consistent with another Rasch-validated questionnaire for use in cataract patients; that is, the ADVS.[19] The changing indications for cataract surgery are well documented,[49] and cataract surgery is now offered at a lower level of impairment in this part of the world than when these instruments were developed, which may explain the unsuitability of the items for the present cataract population. Targeting is sample-dependent[21]; therefore, the Cataract TyPE Spec questionnaire may show optimum targeting in a more severely visually impaired population in regions such as China and Africa.[50] With the exception of the Catquest-9SF,[21] poor targeting has been common to all the Rasch-analyzed visual function questionnaires in cataract patients.[19,34,51] The reason for good targeting using the Catquest-9SF questionnaire was related to optimum targeting of its long form in the same population.

In addition to evenly spacing item difficulties, it is desirable that a questionnaire measures a long span of difficulties. It is relatively easy to capture the functional level of persons who are severely disabled (eg, unable to read a magazine or unable to do usual daily activities); however, it is much more difficult to measure items at the other end of the spectrum. That is perhaps the reason that floor or ceiling effects are commonly seen in the visual function questionnaires. The only solution to this problem is to add more difficult items to suit less impaired patients awaiting cataract surgery, particularly those waiting for second-eye surgery.

Only 1 item (nighttime driving) consistently did not fit the model for the full version and the glare subscale, indicating that the response to this item was influenced by attributes other than patient ability. This misfit was predictable and is in line with findings in other studies, which report that driving items misfit with other items, such as reading.[52,53] In the present study, the misfit was probably because only little more than half the patients (52.4%) drove and most reported greater difficulty with this activity. Deletion of this item confirmed the remaining 11 items fit the Rasch model; however, it did not significantly reduce person separation. This suggests that the item was not congruent to the overall trait, resulting in as much noise as signal, because the removal of a good-fitting item would cause a loss of person separation.[47]

The reduced 11-item version of Cataract TyPE Spec questionnaire represents a unidimensional construct of visual disability in cataract patients. Because the Rasch model is sample-independent, the principal components analysis of the residuals provides a robust demonstration of its dimensionality. This important measurement characteristic reinforces that all the items on the 11-item Cataract TyPE Spec questionnaire hang together as a single construct of visual disability. However, similar unidimensionality was apparent only in 2 subscales: glare and near vision. These subscales can reasonably discriminate between 3 strata of the participant's visual ability. Therefore, investigators can make use of the 2 valid subscales in addition to the overall score or simply implement them individually. Unfortunately, the other subscales of the Cataract TyPE Spec questionnaire did not perform well, as evidenced by their poor person separation. This indicates that these subscales do not meaningfully stratify participants; therefore, we do not endorse their reporting. Dysfunctional subscales have been found in another Rasch revalidated questionnaire too, the original Catquest, and these were removed from the Rasch-scaled revision, the Catquest-9SF.[21]

The Cataract TyPE Spec was free of large DIF in our population, indicating that it is consistent across subpopulations. It showed criterion validity when compared with the self-rating of vision.

We recommend the users of the Cataract TyPE Spec perform Rasch analysis on their own data as populations may vary. However, there may be clinicians and researchers who wish to use the benefits of the Rasch-scaled scores without performing Rasch analysis. Therefore, as stated in our aims, we developed a ready-to-use Microsoft Excel spreadsheet for conversion of raw to Rasch-scaled scores for the 11-item Cataract TyPE Spec instrument (see Supplementary Material which can be downloaded from the journal's web site or obtained by contacting the corresponding author). This spreadsheet converts ordinal category responses into 55 item–category calibrations to create Rasch measurement estimates. We caution that these conversions can be applied only if the sample it is being applied to is similar to that in the present study.

In conclusion, this study confirms the validity and reliability of the Cataract TyPE Spec within a Rasch model. The important limitations of the instrument are that only 2 of the 5 subscales are valid and the instrument poorly targets the population. Poor targeting limits the prospect for measuring the outcome of cataract surgery; thus, other instruments, such as the Catquest-9SF, may be a better choice. Better still would be to develop item banking and computer adaptive testing that can tailor item difficulty to suit the abilities of a given patient. Similar approaches have been used for other areas of health care[54,55] but have yet to be developed for ophthalmology.

REFERENCES

1. McCarty CA, Mukesh BN, Fu CL, Taylor HR. The epidemiology of cataract in Australia. Am J Ophthalmol 1999; 128:446–465
2. Keeffe JE, Taylor HR. Cataract surgery in Australia. Aust N Z J Ophthalmol 1996; 24:313–317
3. Taylor HR. Cataract: how much surgery do we have to do? [editorial] Br J Ophthalmol 2000; 84:1–2. Available at: http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1723236&blobtype=pdf. Accessed May 24, 2009
4. Desai P, Reidy A, Minassian DC, Vafidis G, Bolger J. Gains from cataract surgery: visual function and quality of life. Br J Ophthalmol 1996; 80:868–873. Available at: http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=505640&blobtype=pdf. Accessed May 24, 2009
5. Mönestam E, Wachtmeister L. Dissatisfaction with cataract surgery in relation to visual results in a population-based study in Sweden. J Cataract Refract Surg 1999; 25:1127–1134
6. Brenner MH, Curbow B, Javitt JC, Legro MW, Sommer A. Vision change and quality of life in the elderly; response to cataract surgery and treatment of other chronic ocular conditions. Arch Ophthalmol 1993; 111:680–685. Available at: http://archopht.ama-assn.org/cgi/reprint/111/5/680. Accessed May 24, 2009
7. Mangione CM, Phillips R, 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
8. 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. Available at: http://archopht.ama-assn.org/cgi/reprint/113/10/1305. Accessed May 24, 2009
9. Pesudovs K, Coster DJ. An instrument for assessment of subjective visual disability in cataract patients. Br J Ophthalmol 1998; 82:617–624. Available at: http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1722619&blobtype=pdf. Accessed May 24, 2009
10. 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. Available at: http://archopht.ama-assn.org/cgi/reprint/112/5/630. Accessed May 24, 2009
11. 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
12. Javitt JC, Jacobson G, Schiffman RM. Validity and reliability of the Cataract TyPE Spec: an instrument for measuring outcomes of cataract extraction. Am J Ophthalmol 2003; 136:285–290
13. Lawrence DJ, Brogan C, Benjamin L, Pickard D, Stewart-Brown S. Measuring the effectiveness of cataract surgery: the reliability and validity of a visual function outcomes instrument. Br J Ophthalmol 1999; 83:66–70; Available: http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1722791&blobtype=pdf. Accessed May 24, 2009
14. Massof RW, Rubin GS. Visual function assessment questionnaires. Surv Ophthalmol 2001; 45:531–548
15. Massof RW. The measurement of vision disability. Optom Vis Sci 2002; 79:516–552. Available at: http://www.optvissci.com/pt/re/ovs/pdfhandler.00006324-200208000-00015.pdf%3Bjsessionid=KZYSBB5kbwjgJ6YYXzpgLywwBK16QcpSb8q3bQGC31s8vY3cdGzj!-701738752!181195629!8091!-1. Accessed May 24, 2009
16. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, Denmark, Institute of Educational Research, 1960
17. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. Optom Vis Sci 2007; 84:664–675. Available at: http://www.optvissci.com/pt/re/ovs/pdfhandler.00006324-200708000-00005.pdf;jsessionid=KZcJrBxMKHzzcx4VGhCDgGwV2vDSD5HknjcG729LKxDlX1XPvfm1!-701738752!181195629!8091!-1. Accessed May 24, 2009
18. Mallinson T. Why measurement matters for measuring patient vision outcomes. Optom Vis Sci 2007; 84. Available at: http://www.optvissci.com/pt/re/ovs/pdfhandler.00006324-200708000-00006.pdf;jsessionid=KZcW21QvZ9vvSbXT2Y5S7dNchsSTIwjP2TbTJFyJJYG2tgLLc2LR!-701738752!181195629!8091!-1. Accessed May 24, 2009
19. 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 May 24, 2009
20. Velozo CA, Lai J-S, 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
21. Lundström M, Pesudovs K. Catquest-9SF patient outcomes questionnaire; nine-item short-form Rasch-scaled revision of the Catquest questionnaire. J Cataract Refract Surg 2009; 35:504–513
22. 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
23. Pesudovs K. Patient-centred measurement in ophthalmology – a paradigm shift [commentary]. BMC Ophthalmol 2006; 6:25. Available at: http://www.biomedcentral.com/content/pdf/1471-2415-6-25.pdf. Accessed May 24, 2009
24. 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
25. 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. Available at: http://www.optvissci.com/pt/re/ovs/pdfhandler.00006324-200410000-00009.pdf;jsessionid=KZIQrwyhmlzpQ0JgPykLn0wJq9FITyYqrNBBQCG2kTPPcB41qvh1!-1775402713!181195628!8091!-1. Accessed May 24, 2009
26. Elliott DB, Hurst MA, Weatherill J. Comparing clinical tests of visual function in cataract with the patient's perceived visual disability. Eye 1990; 4:712–717
27. Rubin GS, Muñoz B, Bandeen-Roche K, West SK. Monocular versus binocular visual acuity as measures of vision impairment and predictors of visual disability; for the SEE Project Team. Invest Ophthalmol Vis Sci 2000; 41:3327–3334. Available at: http://www.iovs.org/cgi/reprint/41/11/3327. Accessed May 24, 2009
28. Wright BD, Masters GN. Rating Scale Analysis (Rasch Measurement Book). Chicago, IL, MESA Press, 1982
29. Linacre JM, Wright BD. A User's Guide to WINSTEPS: Rasch model computer program. Chicago, IL, MESA Press, 2000

30. Linacre JM. A User's Guide to WINSTEPS/MINISTEPS; Rasch-Model Computer Programs. Chicago, IL, MESA Press, 2009;. Available at: http://www.winsteps.com/winman/index.-htm?copyright.htm. Accessed May 24, 2009
31. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978; 43:561–573
32. 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 May 24, 2009
33. 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
34. Velozo CA, Lai J-S, 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
35. Comins J, Brodersen J, Krogsgaard M, Beyer N. Rasch analysis of the Knee Injury and Osteoarthritis Outcome Score (KOOS): a statistical re-evaluation. Scand J Med Sci Sports 2008; 18:336–345
36. Wolfe F, Michaud K, Pincus T. Development and validation of the Health Assessment Questionnaire II; a revised version of the Health Assessment Questionnaire. Arthritis Rheum 2004; 50:3296–3305. Available at: http://www3.interscience.wiley.com/cgi-bin/fulltext/109677229/PDFSTART. Accessed May 24, 2009
37. Smith EV Jr. Evidence for the reliability of measures and validity of measure interpretation: a Rasch measurement perspective. J Appl Meas 2001; 2:281–311
38. Wright BD, Linacre JM. Reasonable mean-square fit values. Rasch Measurement Transactions 1994; 8:370. Available at: http://www.rasch.org/rmt/rmt83b.htm. Accessed May 24, 2009
39. Smith RM. Person fit in the Rasch model. Educ Psychol Measurement 1986; 46:359–372
40. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3:205–231
41. Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. London, UK, Lawrence Erlbaum Associates, 2001
42. Linacre JM. A User's Guide to WINSTEPS. Chicago, IL, MESA Press, 2009
43. Teresi JA. Different approaches to differential item functioning in health applications. advantages, disadvantages and some neglected topics. Med Care 2006; 44:S152–S170
44. Holland PW, Wainer H, eds, Differential Item Functioning. Hillsdale, NJ, Lawrence Erlbaum Associates, 1993
45. 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
46. 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
47. 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
48. Lamoureux EL, Pesudovs K, Pallant JF, Rees G, Hassell JB, Caudle LE, Keeffe JE. An evaluation of the 10-item vision core measure 1 (VCM1) scale (the Core Module of the Vision-Related Quality of Life scale) using Rasch analysis. Ophthalmic Epidemiol 2008; 15:224–233
49. 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
50. Foster A. Vision 2020: the cataract challenge [editorial]. J Comm Eye Health 2000; 13(34):17–19. Available at: http://www.cehjournal.org/download/ceh_13_34_017.pdf. Accessed May 24, 2009
51. Pesudovs K, Caudle LE, Rees G, Lamoureux EL. Validity of a visual impairment questionnaire in measuring cataract surgery outcomes. J Cataract Refract Surg 2008; 34:925–933
52. Globe D, Varma R, Azen SP, Paz S, Yu E, Preston-Martin SLos Angeles Latino Eye Study Group. Psychometric performance of the NEI VFQ-25 in visually normal Latinos: the Los Angeles Latino eye study. Invest Ophthalmol Vis Sci 2003; 44:1470–1478. Available at: http://www.iovs.org/cgi/reprint/44/4/1470. Accessed May 24, 2009
53. Massof RW, Hsu CT, Baker FH, Barnett GD, Park WL, Deremeik JT, Rainey C, Epstein C. Visual disability variables. II: The difficulty of tasks for a sample of low-vision patients. Arch Phys Med Rehabil 2005; 86:954–967. Available at: http://download.journals.elsevierhealth.com/pdfs/journals/0003-9993/PIIS0003999304013978.pdf. Accessed May 24, 2009
54. Lai J-S, Cella D, Dineen K, Bode R, Von Roenn J, Gershon RC, Shevrin D. An item bank was created to improve the measurement of cancer-related fatigue. J Clin Epidemiol 2005; 58:190–197
55. Walter OB, Becker J, Bjorner JB, Fliege H, Klapp BF, Rose M. Development and evaluation of a computer adaptive test for 'Anxiety' (Anxiety-CAT). Qual Life Res 2007; 16(suppl 1):143–155