Clinical and Experimental Ophthalmology 2009; 37: 785–794 doi: 10.1111/j.1442-9071.2009.02133.x

Original Article

Catquest questionnaire: re-validation in an Australian cataract population

Vijaya K Gothwal PhD,[1,2] Thomas A Wright BPsyc(Hons),[1] Ecosse L Lamoureux PhD,[3–5] Mats Lundström MD PhD[6] and Konrad Pesudovs PhD[1]

[1]NH & MRC Centre for Clinical Eye Research, Department of Ophthalmology, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, South Australia, [3]Centre for Eye Research Australia, The University of Melbourne, Melbourne, Victoria, [4]Vision CRC, Sydney, New South Wales, Australia; [2]Meera and L B Deshpande Centre for Sight Enhancement, L V Prasad Eye Institute, Hyderabad, India; [5]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore; and [6]EyeNet Sweden, Blekinge Hospital, Karlskrona, Sweden

Abstract

Background: The Catquest questionnaire was developed using traditional methodology to enable cataract surgery outcomes assessment in European countries. Recently, it has been validated using Rasch analysis in a Swedish population resulting in the Catquest-9SF. The aim of the present study was to assess the performance of the Catquest and the Catquest-9SF questionnaires using Rasch analysis in Australian cataract patients.

Methods: A total of 217 cataract patients awaiting surgery at Flinders Medical Centre, Adelaide, South Australia self-administered the Catquest questionnaire. This is a 19-item instrument containing frequency, difficulty, symptoms and global rating items. Rasch analysis was undertaken to assess the unidimensionality (whether all the items are measuring a single underlying construct using principal components analysis or PCA), person separation (ability of the questionnaire to distinguish between strata of patient ability) and targeting of item difficulty to person ability.

Results: Similar to the previous validation study, the original Catquest questionnaire required revision because of misfit and multidimensionality necessitating removal of the frequency items. The revised version was similar to the Catquest-9SF although an extra driving item was a valid optional inclusion. The Catquest-9SF performed well in the Australian cohort satisfying all criteria for valid measurement including unidimensionality. However, targeting of item difficulty to person ability was marginally worse compared with the Swedish cataract population indicating the Australian cataract patients present for surgery at lower levels of visual disability.

Conclusions: The Catquest-9SF is a reliable and valid measure of visual disability in the Australian cataract population.

Key words: Australia, cataract, Catquest, questionnaire, Rasch analysis, reliability.

Introduction

The Catquest questionnaire is a visual disability questionnaire widely used in cataract patients in Europe.[1,2] The questionnaire has been developed and validated in Sweden using traditional classical test theory[3] and has been used in broad variety of studies, primarily cataract outcomes research.[4,5]

■ **Correspondence:** Dr Konrad Pesudovs, NH & MRC Centre for Clinical Eye Research, Department of Ophthalmology, Flinders Medical Centre, Bedford Park, SA 5042, Australia. Email: konrad.pesudovs@flinders.edu.au

Received 23 March 2009; accepted 10 August 2009.

Funding/Support: This study was supported in part by National Health and Medical Research Council (Canberra, Australia) Centre of Clinical Research Excellence Grant 264620. KP is supported by National Health and Medical Research Council (Canberra, Australia) Career Development Award 426765.

However, for a more in-depth examination of basic measurement properties, outcomes measures in ophthalmology are increasingly being investigated using Rasch models.[6–9] Rasch analysis provides psychometric information that is not obtainable through classical test theory.[10,11] Specifically, Rasch analysis enables examination of the functioning of rating scale categories; the validity of a measure (i.e. does the questionnaire measure what it purports to measure) by evaluating the fit of individual items to the underlying construct, and determining whether the items measure a unidimensional construct.[12,13] These attributes are required to justify summation of scores. More recently, Lundström and Pesudovs applied the Rasch model to the Catquest questionnaire on Swedish cataract patients to assess multiple psychometric characteristics: (i) unidimensionality or fit (the extent to which items measure a single construct); (ii) targeting (the extent to which items are of appropriate difficulty for the sample); (iii) item difficulty (the ordering of items from least to most difficult to perform); and (iv) separation (the extent to which the items discriminate distinct levels [strata] of functioning within the domain).[14] However, as a majority of the participants never drove, the two driving items were excluded from the original 19 items for this analysis. The authors found that only the visual disability subscale formed a valid measurement scale and the global rating items could contribute to this measurement.[14] The symptoms and frequency of performing activities items did not form valid subscales and did not contribute to the measurement. The authors proposed refinements of the Catquest questionnaire leading to the creation of its short form, the Catquest-9SF with adequate psychometric properties. Furthermore the Catquest-9SF questionnaire was shown to be highly responsive to cataract surgery and was moderately correlated with visual acuity.[14]

The overall high performance of the Catquest-9SF questionnaire in its reliability, validity and responsiveness to change associated with surgical treatment dictates that its popularity will be further enhanced among the ophthalmic community.[2] Although Catquest-9SF questionnaire had good targeting in the Swedish population, some concerns have been raised about the targeting in other parts of the developed world. Targeting is sample-dependent and therefore makes it important that the Catquest-9SF is tested in other cataract populations. Furthermore, information about the performance of the Catquest questionnaire in different populations is scarce and in general, there is a need to explore the performance of visual disability questionnaires in populations with different clinical and demographic characteristics such as visual acuity, age and gender. Therefore the aim of this study was to test whether the Catquest questionnaire would behave well in an Australian cataract population that is different to Sweden culturally, demographically and in cataract severity and the threshold for cataract surgery.

METHODS

Catquest

The Catquest questionnaire was originally developed in Swedish and was translated into English by the team with an effort to provide a common measure in reporting the results of cataract surgery in English-speaking populations such as Australia.[1]

The Catquest questionnaire contains 19 items in four content areas (each representing a subscale): frequency of performing activities (7 items), perceived difficulty in performing daily-life activities (8 items), global questions about difficulties in general and satisfaction with vision (2 items) and cataract symptoms (2 items). The response options are therefore different for each of these subscales. The four response options for the perceived difficulty levels vary from 1 (no difficulty) to 4 (extreme difficulty) with the intermediate categories of 2 and 3 representing some difficulty and much difficulty, respectively. Therefore a higher score represents greater visual disability. The two symptoms items share the same response options as those assessing difficulty. The four response options for satisfaction with vision include 1 (very satisfied) to 4 (very dissatisfied) with intermediate categories of 2 and 3 representing rather satisfied and rather dissatisfied, respectively. The four response options for the frequency of performing the activity include 1 (do the activity frequently and for watching television it is several hours daily) to 4 (do not do the activity) with intermediate categories of 2 and 3 representing 'do the activity more frequently (2–4 times a week or for watching television, it is at least 1 h daily)' and 'do the activity rarely (once a week at the most)', respectively.

There are other items about things such as home help, other diseases and car driving/employment that have been basically included by the developers as demographic variables and not to evaluate the benefits of surgery and so were excluded from this analysis.

Participants

Patients awaiting a cataract extraction procedure in one eye or both eyes at the Flinders Eye Centre, Flinders Medical Centre, Adelaide, South Australia comprised the participants of this study. Inclusion criteria were age 18 years or older, English-speaking and had no severe cognitive impairment. While on the cataract surgery waiting list (patients wait an average of 3–4 months for surgery) patients were mailed the Catquest questionnaire for self-completion and return via a self-addressed envelope. Ethical approval was obtained from the Flinders Clinical Research Ethics Committee, and all patients who agreed to participate signed a consent form. The study was conducted in accordance with the tenets of the Declaration of Helsinki.

A demographic data form was also included with the questionnaire that contained self-report questions regarding age, gender, the duration of cataract, presence/absence of systemic and ocular comorbidities. The information regarding the comorbidities was then verified from the details filled-in by the treating ophthalmologist in the medical records. This study sample appears to be representative of the elderly cataract population in Australia.[15] The clinical characteristics of the patients who completed the Catquest questionnaire are shown in Table 1.

Clinical assessment

Routine clinical assessments were performed prior to cataract extraction. Visual acuity was tested using a computerized system based on LogMAR principles with a screen illumination of 150 cd/m^2. All the assessments were performed binocularly as binocular acuity is thought to be representative of real world ability.[16,17] Thus we used binocular visual acuity in all our analyses.

Table 1. Clinical characteristics of the 217 participants who completed the Catquest questionnaire

Characteristic	*n* (%) or mean ± SD
Age (years)	75.0 ± 8.8
Gender	
Male	91 (41.9)
Female	126 (58.1)
Habitual binocular visual acuity	
Mean ± SD	
LogMAR	0.22 ± 0.20
Snellen	6/9.5^{-1}
Range	
LogMAR	–0.26 to 0.90
Snellen	6/3^{-2} to 6/48
Awaiting second-eye surgery	88 (41.1)
Ocular comorbidity†	
Present	124 (57.1)
Absent	93 (42.9)
Systemic comorbidity‡	
Present	190 (87.6)
Absent	–27 (12.4)

†Includes glaucoma, diabetic retinopathy, age-related macular degeneration etc. ‡Includes diabetes, hypertension, angina etc.

Rasch analysis

Rasch analysis was performed in two phases: Phase I – analysis of the original Catquest questionnaire and Phase II – analysis of the Catquest-9SF questionnaire.

Rasch Analysis[10,18] was performed with the WINSTEPS software (version 3.68, Linacre[19]) using the Andrich[20] rating scale model. As there were four different question formats, a 4-Andrich rating scale design was applied. However, each rating scale was analysed separately and if each was valid then the option of combining all the rating scales in a single analysis could be considered. Rasch analysis focuses on the psychometric properties of the item, person and rating scale categories.[21] It allows estimates of item difficulty (i.e. how difficult the items are) and person measures ('person ability', representing the extent to which participants or persons possess the trait being examined) to be made along postulated traits, visual disability in the present case. Two values are used throughout the analysis: logit measures and fit statistics. The logit (or log-odds units) is the natural logarithm of the odds of a participant being successful at a specific task or an item being successfully carried out. Conventionally, 0 logit is ascribed to mean item difficulty. For the person category, logit measures indicate whether one person is more able than another (e.g. does one person have better visual ability than another?); for items, logit measures indicate whether one item is more difficult than another (e.g. is reading newspaper print more difficult than recognizing faces of people?). The procedures adopted in this study are consistent with those conducted in the earlier Rasch validation of the Catquest questionnaire in a Swedish sample.[14] Rasch measurement models have been described in detail by Massof.[22]

For a good fitting model, we would expect that, for each of the items, participants with greater visual disability would choose higher categories (such as three or four), and those with low levels of visual disability would consistently choose lower categories (such as one or two). In Rasch analysis terms, this would be indicated by an ordered set of response thresholds for each of the items. If we consider the categories to lie on a scale, then threshold refers to the point of intersection between two adjacent categories where probability of either category being chosen is equal. The number of thresholds for an item is one less than the number of categories. The items in Catquest questionnaire have four categories and therefore have three thresholds. Thus, the first threshold for an item is the ability of participants for whom scoring one and two is equally likely; the so on for second and third thresholds. The thresholds should demonstrate a monotonic (one direction) response process (i.e. one followed by two and so on) that indicates that with increasing disability the probability of selecting higher category for an item would increase in an ordered fashion from least to most difficult. However disordering (for e.g. third threshold being located between first and second) can occur when participants have difficulty differentiating between categories. In such situations, reorganization of categories by combining them is often performed.[23]

On account of the Rasch model being probability-based, some amount of deviation of the scores of items can be expected. When an item does not perform as expected, the fit statistics (i.e. the infit mean square statistic, infit MnSq or simply infit) flag unexpected behaviour of an item.[11] The ideal value of the infit MnSq is 1.0.[10] Items have high infit statistics when they do not measure the same construct as the other items in the set. Items with infit MnSq values between 0.7 and 1.3 were considered acceptable and values outside this range indicated misfitting items that were considered for deletion.[24]

Recent studies suggested that fit statistics (described in the aforementioned) alone are inadequate for determining unidimensionality.[25–29] Therefore, principal components analysis (PCA) of the residuals was also used in combination with Rasch fit statistics to test the unidimensionality of the Catquest.[26–28,30,31] The PCA transforms correlated items into principal components, and the following rules of thumb were used to confirm unidimensionality: a high level of variance such as 60% or greater accounted for by the principal component is indicative of a low likelihood of additional components.[32] Also, if the variance explained by the principal component for the empirical data and model are comparable, it also indicates that there is a low possibility of finding additional components. The first contrast in the residuals indicates whether there are any patterns within the variance unexplained by the principal component to suggest that a second construct is being measured. We used the criterion of an eigenvalue of >2.0 for the first contrast, which indicates that the contrast has the strength of at least two items (this is sufficient evidence of a second construct), as this is greater than the magnitude seen with random data.[32]

An additional feature of Rasch models is that they allow the researcher to examine the items for 'differential item functioning (DIF)', when comparing one subgroup with another.[33–35] Within the framework of Rasch measurement, the questionnaire should work in the same way, regardless of the group being assessed. Thus, in the case of visual disability, the probability of a participant affirming an item (or category) at a given level of disability should be the same for men or women, younger or older participants and so on. Items that do not yield the same item response function for two or more subgroups display DIF. We selected the DIF variables a priori in our study. We examined all items for the extent to which their functioning was differentially affected by age (<75 and ≥75 years; 75 being the median age), gender, cataract status (first eye *vs*. second eye surgery), presence or absence of systemic and ocular comorbidity. DIF was considered absent if it was less than 0.50 logit, minimal but probably inconsequential if it ranged between 0.50 and 1.0 logit and notable if it was >1.0 logits.[36,37]

The overall measurement precision of the Catquest questionnaire was assessed by examining the person separation statistics (expressed in logits). Person separation indicates the number of distinct strata of persons that can be discerned by the questionnaire.[38] The larger the person separation, the greater the number of distinct levels of functioning that can be distinguished by the questionnaire. A person separation of 2.00 represents the minimum acceptable level of separation (indicative of three discernable strata), and 3.00 represents an excellent level of separation.[10]

The Rasch model describes the hierarchy of items from most to least commonly endorsed. The trait, that is visual disability, described by this hierarchy should conceptually make sense or conform to the theoretical order described by the developers or on the basis of clinical experience. Items that are commonly endorsed should be those characteristics that are more commonly observed or statements that even participants possessing little of the trait would endorse. Conversely, items that are less commonly endorsed should be those characteristics that are less commonly observed or statements that only participants possessing a great deal of the trait would endorse. The hierarchical order of the Catquest items was examined using the person-item map provided by the WINSTEPS software. Such item hierarchy enables comparison of item difficulty with person ability and can be used to determine whether the items of the Catquest questionnaire cover the range of person abilities in the sample (i.e. reveal ceiling or floor effects).[39] The average person measure was used to determine the extent to which a set of items is of appropriate difficulty for the participants. An absolute average person measure ≥0.5 logit indicates mistargeting (i.e. mismatch between abilities of participants and difficulty posed by the items).[40]

The short-form Catquest 9-SF and the four subscales were also analysed similar to the full version of the Catquest questionnaire.

Finally, criterion validity was assessed by determining the relationship binocular visual acuity and Rasch-scaled Catquest scores.

Descriptive statistics were analysed using SPSS software version 15.0 (SPSS Inc., Chicago, IL, USA), and statistical significance was set at $P < 0.05$.

Results

Phase I – Assessment of the native Catquest questionnaire

Behaviour of the rating scale

Category thresholds were ordered indicating that the participants were using the rating scale as was intended by the developers.[5]

Overall performance

The person separation reliability was satisfactory indicating that the Catquest questionnaire was able to reliably distinguish among several groups of participants (Table 2).

Unidimensionality

Four items (21.0%) misfit (Table 2). PCA of the residuals showed that the variance explained by the measures was comparable for the empirical calculation (56.1%) and by the model (53.8%). The unexplained variance explained by the first contrast was 2.3 eigenvalue units, and there were no further contrasts with eigenvalue >2.0. Taken together these findings indicated that the Catquest questionnaire was not unidimensional. Six items loaded (>0.3) positively on to the first contrast and the examination of the content of these revealed them to be all related to assessing the frequency of performing an activity. We examined if these items could form a separate subscale, but they lacked person separation and so were not a valid subscale.

Differential item functioning

Five items showed minimal DIF by the demographic variables assessed. However two items showed notable DIF.

Subscales

None of the subscales had adequate person separation indicating they were unable to distinguish strata of participants (Table 3).

Item reduction

The Catquest questionnaire is not unidimensional and contains a secondary dimension caused by the frequency items. Further support for this lack of unidimensionality came from misfitting items indicating that they were not contributing towards the measurement of the underlying construct (i.e. visual

Table 2. Overall performance of the Catquest, Catquest-9SF in 217 Australian cataract patients and comparison with previous Rasch analysis of the Catquest-9SF in Swedish cataract patients[14]

Versions	Original Catquest	Catquest-9SF		Catquest-9SF plus driving item	Catquest-9SF plus driving and symptoms items
		Present study (Australian)	Previous study (Sweden)		
No. of items	19	9	9	10	12
No. of misfitting items	4	0	0	0	0
Person separation	2.42	2.28	2.65	2.31	2.73
Mean item location	0	0	0	0	0
Mean person location	–0.83	–0.86	–0.34	–0.96	–0.80
Principal components analysis (eigenvalue for first contrast)	2.3	1.7	1.6	1.8	1.8

Table 3. Overall performance of subscales of the Catquest questionnaire

Parameters	Subscales			
	Frequency of daily-life activities	Difficulty in performing daily-life activities	Global assessment	Cataract symptoms
No. of items	7	8	2	2
No. of misfitting items	2	0	0	0
Person separation	0.90	1.94	1.30	0
Mean item location	0	0	0	0
Mean person location	–1.09	–0.88	0.65	–1.67

Table 4. Fit statistics for the Catquest-9SF, 10 and 12 items to the Rasch model

Item description	Infit mean square		
	Catquest-9SF	Catquest-9SF plus driving item	Catquest-9SF plus driving and symptoms items
Read newspaper	0.76	0.76	0.72
Recognize faces	1.22	1.24	1.23
Prices when shopping	0.86	0.87	0.88
Walk on uneven ground	0.98	0.99	0.98
Needlework	0.91	0.92	0.92
Seeing text on TV	1.21	1.23	1.25
Hobbies	1.01	0.98	0.94
Daily-life activities in general	0.86	0.86	0.84
Satisfaction with vision	1.19	1.16	1.11
Difficulty while driving	NA	0.97	1.22
Visual disturbances	NA	NA	0.93
Headlights, lamps, sunlight reduce vision	NA	NA	0.94

Infit mean square, information weighted mean square fit statistic, all items are within the desired range (0.7–1.3); NA, not applicable.

disability). The items from the secondary dimension (six frequency items) were best eliminated, which left a set of 13 items in which one item (item no. 4 from the original 19 items) continued to misfit (last remaining frequency item). Following its deletion, there were 12 items (9 items identical to the Catquest-9SF, 2 symptoms items and 1 driving item) in the core questionnaire (Table 4). Because all items fit the construct, this suggests that a 12-item version of Catquest is a valid measurement instrument. This instrument includes visual disability, global rating, driving and symptoms items. The major difference is the inclusion of items from the symptoms domain that did not fit in the Swedish Rasch re-validation of the Catquest. To be more in line with the Catquest-9SF, we could reduce this to the same nine items with an optional 10th item being driving.

Phase II – Assessment of the Catquest-9SF questionnaire

Overall performance

The person separation was adequate indicating that the Catquest-9SF could differentiate between several strata of participant ability (Table 2).

Unidimensionality

None of the items misfit indicating that they were all contributing to the measurement of the underlying construct (Table 2). PCA of the residuals showed that the variance explained by the measures was comparable for the empirical calculation (65.0%) and by the model (63.7%). The unexplained variance explained by the first contrast was 1.7 eigenvalue units. Taken together these findings indicated that the Catquest-9SF was unidimensional.

Differential item functioning

Two items showed minimal DIF. Participants awaiting cataract surgery in their first eye were more satisfied (0.86 logit) with their vision than those awaiting surgery in the second eye. Similarly, women rated the item 'seeing to walk on uneven ground' 0.50 logit easier as compared with men.

Item hierarchy and targeting

Figure 1 shows distribution of participants' ability and item difficulty measures of the Catquest-9SF. There was a fairly even spread of items along the variable and the participant ability demonstrated a 15.5 logits spread (−8.3 to 7.2 logits; mean = −0.84); the levels of visual disability were on average lower than the mean difficulty of the Catquest-9SF items (set by convention at 0 logit). Item difficulty demonstrated a 5.11-logit spread (−2.94 to 2.17 logits), and the items showed slight mistargeting indicating that the items were easy for the visual abilities of the participants. The most difficult item was 'satisfaction with one's present vision' and easiest was 'recognizing faces of people'.

Therefore, the Catquest-9SF questionnaire showed good person separation, slight mistargeting and unidimensionality and was free of any large DIF.

Comparison of the Catquest-9SF and 10 and 12 item reduced versions

Table 4 provides the fit statistics for each item included in the Catquest-9SF, 10 (adding the

```
            PERSONS - MAP - ITEMS
      <more disabled>|<easier>
   7               . +
                     |
                     |
                     |
   6                 +
                     |
                     |
                   # |
   5                 +
                     |
                     |
                     |
   4               . +
                     |
                    T|
                   . |
   3                 +
                   # |T
                   # |
                   . |  recognizing faces_Q8
   2              .# +
                   # |  daily-life activities in general_Q14
                #### |S
                .## S|
   1            #### +
                #### |
            .####### |  walk on uneven ground_Q10; hobbies_Q13
              .##### |
   0            #### +M
                #### |  reading newspaper_Q7;prices when shopping_Q9
           .######## |
           #########M|  seeing text on TV_Q12
  -1         .###### +  needlework and handicraft_Q11
               ##### |
                #### |S
                 .## |
  -2               # +
                   # |
               .#### |
                  ## |T
  -3              ##S+  satisfaction with vision_Q15
                   . |
                 .## |
                   . |
  -4              .# +
                 .## |
                     |
                     |
  -5                 +
                    T|
                     |
                   # |
  -6          .##### +
                     |
                     |
                     |
  -7                 +
                     |
                     |
                     |
  -8            #### +
      <less disabled>|<more difficult>
```

Fi gur e

Figure 1. Person-item map of the Catquest-9SF. The participants (represented by the # symbol) are on the left of the dashed line with less disabled participants located at the bottom of the map. Items are located on the right of the dashed line, and more difficult items are also located at the bottom of the map. Each '#' and each '.' represents two and one participant, respectively. M, mean; S, 1 SD from the mean; T, 2 SD from the mean.

driving item) and 12 item (adding the symptoms items) versions. All items fit the Rasch model in each of the versions as evidenced by acceptable infit statistics suggesting that the items were contributing to the measurement of the underlying construct. As described in the aforementioned, the Catquest-9SF has adequate person separation reliability.

The addition of single driving item (visual disability) to Catquest-9SF improved the person separation, albeit marginally. However, targeting worsened further (to −0.96 logit).

In comparison, further addition of two symptoms items (visual disability) to the 10 items (Catquest-9SF plus driving item) increased the person separation by a large amount (0.45) and improved targeting marginally (by 0.06 logit). However slight mistargeting persisted.

Criterion validity

Binocular visual acuity (habitual) showed a fair, statistically significant correlation with the Rasch scaled Catquest-9SF score ($r = 0.21$, $P = 0.002$).

Discussion

In its original form, the Catquest questionnaire was not unidimensional as was evidenced by large numbers of misfitting items (all those assessing frequency of performance of an activity). This was further supported by PCA of the residuals. The finding that the native Catquest questionnaire lacked fit to the Rasch model substantiated the findings of the previous Rasch analysis of the questionnaire.[14] However the frequency of performing an activity was included in the original development process with the reasoning that the importance of an activity could be gauged by the frequency of its performance.[1] Nevertheless both the studies bear out that people acknowledge the frequency or the importance of an activity when they respond to disability questions making the information from asking the frequency of an activity redundant. Misfitting items such as the frequency items in the Catquest questionnaire induce measurement noise and do not contribute to the underlying construct; rather are influenced by attributes other than the underlying construct.[41–43] As noted in the methods, such misfitting items are best deleted[44] and following deletion, the Catquest questionnaire had 12 items that comprised of disability (7 items), global rating (2 items), symptoms (2 items) and driving (1 item). Although the disability aspect of the items was similar to the Catquest-9SF, the symptoms and the driving item fit the Rasch model in the Australian cataract patients. However, given the fact that the Catquest questionnaire has only been recently re-validated using Rasch analysis (although in a different population), we considered it appropriate to use the pre-existing Catquest-9SF so as be able to draw suitable comparisons. Additionally as mentioned in original development study of the Catquest questionnaire,[1] the evaluation of the questionnaire was made using a decision tree, and the most important part was the seven disability items that formed the core aspect of the Catquest-9SF[14] and is consistent with the results of the present study.

The analysis then focused on the performance of the Catquest-9SF. The questionnaire behaved well and it appears to be a short, reliable, valid and unidimensional measure to assess the visual disability in Australian cataract patients. However the targeting of item difficulty to patient ability was better in Swedish cataract patients as compared with the Australian cohort suggesting that the items were more suited to assess the visual difficulties of the former (where it was developed) than the latter. This occurs because targeting is sample-dependent, as speculated in the earlier study. As noted in the methods, our participants completed the Catquest questionnaire preoperatively only and at present we do not have postoperative data. Despite the absence of this data, one could speculate that the poor targeting observed preoperatively would only worsen postoperatively and one could foresee a ceiling effect as a result of improved visual functioning. The relatively lower difficulty of items for our participants preoperatively does not indicate that they did not have any visual disability because all the participants were drawn from the cataract surgical waiting list and so by definition all suffered visual disability resulting from cataract[45–47] as has been shown previously.[15] Nonetheless, the profile of the participants in the present study is likely to be representative of cataract populations at public hospitals in any capital city of Australia. The mean age and gender distribution being comparable in the present and the Swedish study, one of the reasons for the mistargeting in the present study is perhaps related to the higher mean visual acuity (0.22 logMAR or 6/9.5−) of the participants in this group. Furthermore the threshold for cataract surgery in Australia has lowered over the recent years.[48]

There was no large DIF for any of the items in the present study indicating that the items were behaving similarly across different subgroups of participants. However, two items showed minimal DIF. Of these, the item 'seeing to walk on uneven ground', consistent with the previous Rasch analysis, was more likely to be endorsed by women. Although the removal of this item would have resulted in a small reduction in the person separation (from 2.28 to 2.15), it was still retained in the questionnaire, as the DIF was minimal. One possible explanation for the DIF for this item is frequency of performance of this activity – *walk on uneven ground*. Traditionally, more men than women (62% *vs*. 38%) drove and so it is possible that men did not do much of outdoor walking as much as women and therefore women came across more uneven ground than men.

The Catquest-9SF showed excellent precision in the Australian cohort indicating that it can reliably distinguish between several groups of participants and are sufficiently reliable for individual patient use.[11] This finding is similar to the previous Rasch validation study of the Catquest questionnaire.[14] Measurement precision is determined by the number of units into which the range is divided and is defined mainly by the number of items in the questionnaire. It is difficult to have satisfactory measurement precision with a small number of items.[39] Attempts to create short-form questionnaires from other questionnaires have resulted in minimum item sets of 12 items for the visual disability assessment, 16 items for the activities of daily vision scale and 10 items for the VF-14.[6,9,49] However, both the ADVS and VF-14 showed poor targeting and the VF-10 required additional new items to optimize measurement.[6,9,49] In comparison, the excellent precision of the Catquest-9SF is appropriate for most research purposes and so makes it a good choice for assessment of visual disability in cataract patients in future studies.

In contrast to the previous Rasch analysis of the Catquest questionnaire[14] and research in quality of life instruments where symptoms and disability failed to tap the same latent trait,[43,50] the symptoms items, however, fit the Rasch model of the Catquest questionnaire in the Australian cataract patients. This indicates that the symptoms items are also contributing to the measurement of visual disability in the cataract patient. One could argue a loss of face validity, but the data clearly show the same latent trait is being tapped so the symptoms items can be retained, although only in this population. The addition of the two symptoms items improved the person separation (2.73) as well as targeting (–0.80 logit) of the Catquest-9SF (making it 12 items). However the addition of the single driving item to the Catquest-9SF did not alter the person separation (2.31) much but worsened the targeting (–0.96 logit). Based on these findings, if a user is looking for higher precision, he/she could consider using the two additional symptoms that would help distinguish relatively greater number of groups of participants. Otherwise, the user of Catquest should revert to the Catquest-9SF that has been shown to be valid in two populations now. For the Australian, or similar, population an optional additional driving item could be included.

Although the actual item difficulty estimates of the Catquest-9SF in this study were different from the previous Rasch analysis of this questionnaire, the diffculty hierarchy (rank order) was quite similar between the two studies.[14] Such visualization of the ordering of items is possible with Rasch analysis but is not available with the classical test theory. This finding provides support for comparability of the Catquest-9SF questionnaire across different cataract populations and the robustness of the present findings.

The Catquest-9SF[1] represents the state of the art for the measurement of visual disability, likely superior to the Rasch-analysed versions of the ADVS[6] and the VF-14[9] and certainly superior to any non-Rasch-analysed visual disability instruments. In conclusion, Rasch analysis revealed that the Catquest-9SF questionnaire is a robust, unidimensional measure largely free of DIF that worked well in Australian cataract patients.

Acknowledgements

The authors wish to thank all the participants of the study.

References

1. 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**: 1226–36.
2. 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**: 968–74.
3. Lundström M, Fregell G, Sjoblom A. Vision related daily life problems in patients waiting for a cataract extraction. *Br J Ophthalmol* 1994; **78**: 608–11.
4. Lundström M, Brege KG, Floren I, Stenevi U, Thorburn W. Impaired visual function after cataract surgery assessed using the Catquest questionnaire. *J Cataract Refract Surg* 2000; **26**: 101–8.
5. Lundström M, Wendel E. Duration of self assessed benefit of cataract extraction: a long-term study. *Br J Ophthalmol* 2005; **89**: 1017–20.
6. Pesudovs K, Garamendi E, Keeves JP, Elliott DB. The activities of daily vision scale for cataract surgery outcomes: re-evaluating validity with Rasch analysis. *Invest Ophthalmol Vis Sci* 2003; **44**: 2892–9.
7. 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**: 4732–41.
8. Lamoureux EL, Pesudovs K, Pallant JF *et al.* An evaluation of the 10-item vision core measure 1 (VCM1) scale (the Core Module of the Vision-Related Quality of Life scale) using Rasch analysis. *Ophthalmic Epidemiol* 2008; **15**: 224–33.
9. 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; **4**: 667–80.
10. Wright BD, Masters GN. *Rating Scale Analysis*. Chicago, IL: MESA Press, 1982.
11. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. London: Lawrence Erlbaum Associates, 2001.

12. Embertson SE, Reise SP. *Item Response Theory for Psychologists*. Marwah, NJ: Erlbaum Associates, 2000.
13. Hambleton RK, Swaminathan H. *Item Response Theory*. Boston, MS: Kluwer-Nijhoff, 1985.
14. 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**: 504–13.
15. Kirkwood BJ, Pesudovs K, Latimer P, Coster DJ. The efficacy of a nurse-led preoperative cataract assessment and postoperative care clinic. *Med J Aust* 2006; **184**: 278–81.
16. Rubin GS, Bandeen-Roche K, Huang GH *et al.* The association of multiple visual impairments with self-reported visual disability: SEE project. *Invest Ophthalmol Vis Sci* 2001; **42**: 64–72.
17. Elliott DB, Hurst MA, Weatherill J. Comparing clinical tests of visual function in cataract with the patient's perceived visual disability. *Eye* 1990; **4**: 712–17.
18. Wright BD, Stone MH. *Best Test Design*. Chicago, IL: MESA Press, 1979.
19. Linacre JM. WINSTEPS Rasch measurement (computer program). Chicago, IL: Winsteps.com, 2008.
20. Andrich DA. A rating scale formulation for ordered response categories. *Psychometrika* 1978; **43**: 561–73.
21. Pesudovs K, Burr JM, Elliott DB. The development, assessment and selection of questionnaires. *Optom Vis Sci* 2007; **84**: 664–75.
22. Massof RW. The measurement of vision disability. *Optom Vis Sci* 2002; **79**: 516–52.
23. 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**: 925–33.
24. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Measurement Transactions* 1994; **8**: 370.
25. Smith EV Jr, Johnson D. Attention deficit hyperactivity disorder: scaling and standard setting using Rasch measurement. *J Appl Meas* 2000; **1**: 3–24.
26. Linacre JM. Structure in Rasch residuals: why principal components analysis? *Rasch Measurement Transactions* 1998; **12**: 636.
27. Linacre JM. Detecting multidimensionality: which residual data-type works best? *J Outcome Meas* 1998; **2**: 266–83.
28. 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**: 205–31.
29. Smith RM, Miao CY. Assessing unidimensionality for Rasch measurement. In: Wilson M, ed. *Assessing Unidimensionality for Rasch Measurement*. Norwood, NJ: Ablex Publishing Co., 1994; 316–27.
30. Linacre JM. *Learning from Principal Components Analysis of Residuals*. Chicago, IL: COMET, 1999.
31. Moore D, McCabe G. Introduction to the practice of statistics. In: Freeman WH, ed. *Introduction to the Practice of Statistics*, Vol. 92. New York: W.H. Freeman and Co, 1993; 106.
32. Linacre JM. *A User's Guide to WINSTEPS*. Chicago, IL: Winsteps.com, 2005.
33. Holland PW, Wainer H. Differential item functioning. In: Holland PW, Weiner H, eds. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum, 1993.
34. Scheuneman JD. A method for assessing bias in test items. *J Educ Meas* 1979; **16**: 143–52.
35. Smith RM, Suh KK. Rasch fit statistics as a test of the invariance of item parameter estimates. *J Appl Meas* 2003; **4**: 153–63.
36. Wright BD, Douglas GA. Best test design and self-tailored testing. *Research Memorandum No. 19*. Chicago, IL: Statistical Laboratory, Department of Education, University of Chicago, 1975.
37. Wright BD, Douglas GA. Rasch item analysis by hand. *Research Memorandum No. 21*. Chicago, IL: Statistical Laboratory, Department of Education, University of Chicago, 1976.
38. 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.
39. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Med Care* 2004; **42**: I17–24.
40. Prieto L, Alonso J, Lamarca R. Classical test theory vs. Rasch analysis for quality of life questionnaire reduction. *Health Qual Life Outcomes* 2003; **1**: 27.
41. White LJ, Velozo CA. The use of Rasch measurement to improve the Oswestry classification scheme. *Arch Phys Med Rehabil* 2002; **83**: 822–31.
42. Smith RM. Person fit in the Rasch model. *Educ Psychol Meas* 1986; **46**: 359–72.
43. 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.
44. Smith RM, Schumacker RE, Bush MJ. Using item mean squares to evaluate fit to the Rasch model. *J Outcome Meas* 1998; **2**: 66–78.
45. Pesudovs K, Coster DJ. Cataract surgery reduces subjective visual disability. *Aust N Z J Ophthalmol* 1997; **25**(Suppl. 1): S3–5.
46. Pesudovs K, Coster DJ. An instrument for assessment of subjective visual disability in cataract patients. *Br J Ophthalmol* 1998; **82**: 617–24.
47. Lamoureux EL, Hooper CY, Lim L *et al.* Impact of cataract surgery on quality of life in patients with early age-related macular degeneration. *Optom Vis Sci* 2007; **84**: 683–8.
48. Taylor HR, Vu HT, Keeffe JE. Visual acuity thresholds for cataract surgery and the changing Australian population. *Arch Ophthalmol* 2006; **124**: 1750–3.
49. Pesudovs K, Elliott DB, Coster DJ. The Cataract Outcomes Questionnaire – a Rasch scaled measure of visual disability. *Invest Ophthalmol Vis Sci* 2005; **45**: E-Abstract 3844.
50. Pesudovs K, Garamendi E, Elliott DB. The Quality of Life Impact of Refractive Correction (QIRC) Questionnaire: development and validation. *Optom Vis Sci* 2004; **81**: 769–77.