

# The outcome of cataract surgery measured with the Catquest-9SF

Mats Lundstrom,<sup>1</sup> Anders Behndig,<sup>2</sup> Maria Kugelberg,<sup>3</sup>  
Per Montan,<sup>3</sup> Ulf Stenevi<sup>4</sup> and Konrad Pesudovs<sup>5</sup>

<sup>1</sup>EyeNet Sweden, Blekinge Hospital, Karlskrona, Sweden

<sup>2</sup>Department of Ophthalmology, Umeå University Hospital, Umeå, Sweden

<sup>3</sup>St. Erik Eye Hospital, Stockholm, Sweden

<sup>4</sup>Department of Ophthalmology, Sahlgrenska University Hospital, Mölndal, Sweden

<sup>5</sup>NH&MRC Centre for Clinical Eye Research-Flinders University and Flinders Medical Centre, Adelaide, South Australia, Australia

## ABSTRACT.

**Purpose:** The purpose of this study was to use the Catquest-9SF to measure cataract surgery outcomes, and to use Rasch analysis to test the psychometric properties of this questionnaire, including its validity and responsiveness.

**Methods:** Patients were recruited as consecutive cataract surgery patients during 1 month at six surgical units in Sweden (via the National Cataract Register). The patients completed the questionnaire before surgery and 3 months after. The Catquest-9SF data were assessed for fit to the Rasch model using version 3.63.2 of the WINSTEPS software (Winsteps.com, Beaverton, OR, USA). Both preoperative and postoperative questionnaires were included in the analysis. The responsiveness to cataract surgery was calculated as the effect size.

**Results:** Completed questionnaires before and after surgery were received from 846 patients. The Rasch analysis showed that the category thresholds were ordered. All items fit a single overall construct (infit range 0.79–1.40; outfit range 0.74–1.40). The ability to discriminate different strata of person ability was good, with a real patient separation of 2.58 and patient separation reliability of 0.87. The questionnaire showed unidimensionality and was largely free from differential item functioning. The item difficulty was reasonably well targeted to both preoperative and postoperative patient ability. The Catquest-9SF Rasch score correlated significantly with visual acuity, and cataract surgery resulted in a significant improvement with an effect size of 1.8.

**Conclusion:** The Catquest-9SF shows excellent psychometric properties, as demonstrated by Rasch analysis. It is highly responsive to cataract surgery, and its brevity (nine items) makes it well suited for use in daily clinical practice.

**Key words:** cataract surgery – questionnaire – Rasch analysis – response theory – validation – visual disability – visual function

Acta Ophthalmol.

© 2009 The Authors

Journal compilation © 2009 Acta Ophthalmol

doi: 10.1111/j.1755-3768.2009.01801.x

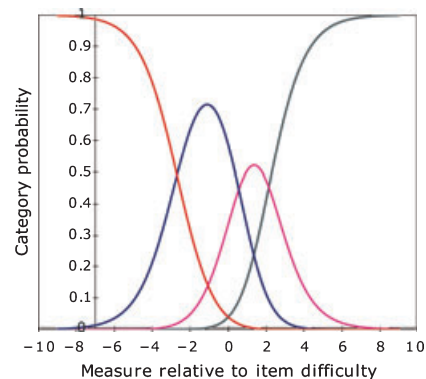
## Introduction

The Catquest questionnaire (Lundström et al. 1997) has been included since 1995 in the follow-up studies after cataract surgery performed by the Swedish National Cataract Register (NCR) (Lundström et al. 2002). The questionnaire is used to measure patients' self-assessed visual function before and 6 months after a cataract extraction to evaluate the benefit of the surgery. The items in Catquest are grouped within four dimensions: daily life activity level, perceived difficulties in performing activities of daily living, cataract symptoms, and global questions about difficulties in general and satisfaction with vision. Each item has four response categories, which are allocated ordinal numerical values by means of a Likert scale. The items within each dimension result in a total score based on the Likert scale, and the overall evaluation is made using a decision tree (Lundström et al. 1998). The most important dimension in this evaluation is perceived difficulties in performing activities of daily living. This method of scoring has limitations, most importantly the incorrect assumption that the method of scoring produces an interval scale. Modern test theory, including Rasch analysis, has demonstrated the invalidity of summary scoring (Wright & Linacre 1989; Fisher et al. 1997;

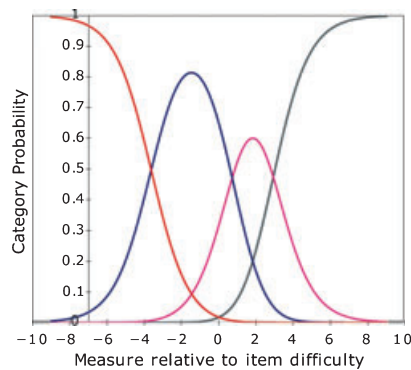
**Table 1.** Patient characteristics.

Patients ( <i>n</i> )	846
Mean age (year) ± SD	75.6 (±9.2)
Gender, <i>n</i> (%)	
Female	539 (63.7)
Male	307 (36.3)
First-eye surgery (%)	58.2
Second-eye surgery (%)	41.8
Vision impairing ocular comorbidity, <i>n</i> (%)	269 (31.8)
Best corrected visual acuity	
Before surgery	
Surgery eye; mean logMAR ± SD	0.46 ± 0.27
Fellow eye; mean logMAR ± SD	0.25 ± 0.25
After surgery	
Surgery eye; mean logMAR ± SD	0.13 ± 0.20

Pesudovs 2006) and shown that the problem can be resolved by providing linear interval transformation of the ordinal raw score, thereby permitting the use of parametric statistical techniques on the questionnaire data (Norquist et al. 2004; Garamendi et al. 2006; Pesudovs 2006). Other unique evaluations available in Rasch analysis include assessment of how well the difficulty of an item targets the level of person ability in the population and measurement of scale validity – in particular, item and person fit to the overall construct (Fisher et al. 1997; Mallinson et al. 2004; Pesudovs et al. 2007). A number of ophthalmic questionnaires have been assessed using Rasch analysis (Gothwal et al. 2009) including the original Catquest questionnaire (Lundström & Pesudovs 2009). From this analysis, a nine-item short form of Catquest was suggested, the Catquest-9SF. Catquest-9SF fulfilled all demands of being a valid measure as evaluated through Rasch analysis. However, this analysis was based on data from nine items selected from the complete list of items in the original Catquest. It is

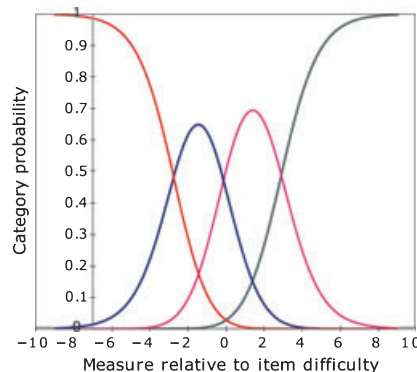


**Fig. 1.** Category probability curve for the item 'reading text in the newspaper'.



**Fig. 2.** Category probability curve for the global question about any difficulties in daily life.

thus essential to investigate whether the suggested short form of Catquest is a valid measure on its own without extracting items from a longer completed questionnaire. In this study, we tested the Catquest-9SF for the first time on cataract patients before and 3 months after surgery. In modern cataract surgery with rapid rehabilitation a 3-month follow-up time is enough. The purpose was to assess the ability of the instrument to measure cataract surgery outcomes,



**Fig. 3.** Category probability curve for the 'satisfaction with vision' item.

including the responsiveness of the questionnaire, and at the same time to assess its psychometric quality and validity using Rasch analysis.

## Patients and Methods

### Catquest-9SF

The Catquest-9SF contains nine items all of which are supposed to measure the same underlying trait – the patient's self-assessed visual disability. Visual disability means the patient's self-assessed restriction and difficulty in having a normal visual function for everyday tasks. Each item has four response categories. For eight items, the categories are as follows: 'no, no problem'; 'yes, some problems'; 'yes, great problems'; and 'yes, very great problems'. For the ninth item, a question about the patients' satisfaction with their overall visual function, the categories are as follows: 'yes, very satisfied'; 'yes, fairly satisfied'; 'no, rather dissatisfied'; and 'no, very dissatisfied'. All of the items include an additional option of 'can't say'; this category is treated as missing data in the following analysis. The items are specified in Table 2. The items are presented in the same format in both the preoperative and postoperative versions of the questionnaire. In this study, we include both preoperative and postoperative data to evaluate the validity of the questionnaire for both situations. All questionnaires were completed in Swedish. This means that the information presented here about question and response format is given in a translation.

### Rasch analysis

The Catquest-9SF data were assessed for fit to the Rasch model (Rasch 1960) using version 3.63.2 of the WINSTEPS software produced by John M Linacre (Linacre 2002, 2006) and the Andrich (Andrich 1978) version of Rasch model estimates based on joint maximum likelihood estimation. A Rasch analysis compares the level of difficulty required to perform a task addressed in the items with the respondent's level of ability to perform that task. Both item difficulty and person ability are sorted on the same linear scale. If the data meet the expectations of the Rasch model, then a transformation of the

**Table 2.** Item fit characteristics.

Item	Item calibration (standard error)	Infit	Outfit	DIF pre-op to post-op
For the seven difficulty items		MNSQ*	MNSQ*	
Do you have difficulty with the following activities because of your vision				
If so: yes – very great difficulties, yes – great difficulties, yes – some difficulties, no – no difficulties				
1. Reading text in the newspaper	–0.15 (0.05)	1.02	1.02	0.17
2. Recognizing faces of people you meet	1.30 (0.05)	1.40	1.40	0.07
3. Seeing prices of goods when shopping	–0.36 (0.05)	0.95	0.99	0.62
4. Seeing to walk on uneven ground	0.20 (0.05)	1.21	1.36	0.52
5. Seeing to do needlework and handicraft	–0.61 (0.05)	0.87	0.84	0.40
6. Reading text on TV	0.05 (0.05)	1.02	1.00	–0.32
7. Seeing to carry out a preferred hobby	0.46 (0.05)	0.86	0.80	0.0
Two global assessment items				
8. Do you experience that your present vision gives you difficulties in any way in your daily life?	0.16 (0.05)	0.79	0.74	–0.47
9. Are you satisfied or dissatisfied with your present vision?	–1.05 (0.05)	0.85	0.86	–0.99

MNSQ\*, weighted mean square statistics.

ordinal raw score into a true Rasch scale is achieved. This scale is linear, and uses a unit known as the logit, which is the natural logarithm of the odds ratio. An important aspect is that the Catquest-9SF should be valid for measurement on both preoperative and postoperative patient data, and so Rasch analysis was performed on preoperative and postoperative data stacked as a single dataset (846 cases but 1692 response sets) (Wolfe & Chiu 1999).

Rasch analysis tests the following psychometric properties of the questionnaire:

- *Rating scale.* This test investigates whether the category thresholds are ordered. Catquest-9SF uses four response categories, and so there are three thresholds between the response probabilities.
- *Ability to discriminate different strata of person ability.* This is a reliability test which results in two quantities: person separation and separation reliability. The separation reliability coefficient represents the precision of the item measures.
- *Item fit statistics.* Each item should contribute to a picture of the respondent's ability in a predictable way. The analysis produces two fit statistics: infit and outfit mean square. Both these fit statistics should have a value of 1, with suggested limits of 0.7 and 1.3.
- *Targeting precision of the instrument to the studied population.* Ideally, the ability of the patients and the difficulty of the questions should centre on the same mean. There should be meaningful items for the more able

patients as well as for the less able patients.

- *Unidimensionality.* This signifies that the items in the instrument measure one and the same underlying trait. Unidimensionality is necessary to create and use a summary score.

- *Differential item functioning (DIF).* This test evaluates whether a specific group of patients (e.g. divided by gender, co-morbidity, cataract status, etc.) responds differently to an individual item despite similar levels of the underlying trait being measured. DIF can cause misfitting of the data to the Rasch model. It can be detected statistically using analysis of variance to compare scores across each level of the person factor and across different levels of trait (referred to as class intervals), or by inspection of the raw differences in item calibration between groups. Testing for DIF was performed between groups differing by gender, ocular co-morbidity, first-eye and second-eye cataract surgery, and preoperative versus postoperative answers.

### Patients

Patients were recruited as consecutive cataract surgery patients during March 2008 at six surgical units in Sweden affiliated to the Swedish NCR. This type of study is repeated by the NCR once a year during the month of March as a routine part of follow-up evaluation. In previous years, the original Catquest was used. The questionnaire (Catquest) and the method were 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.

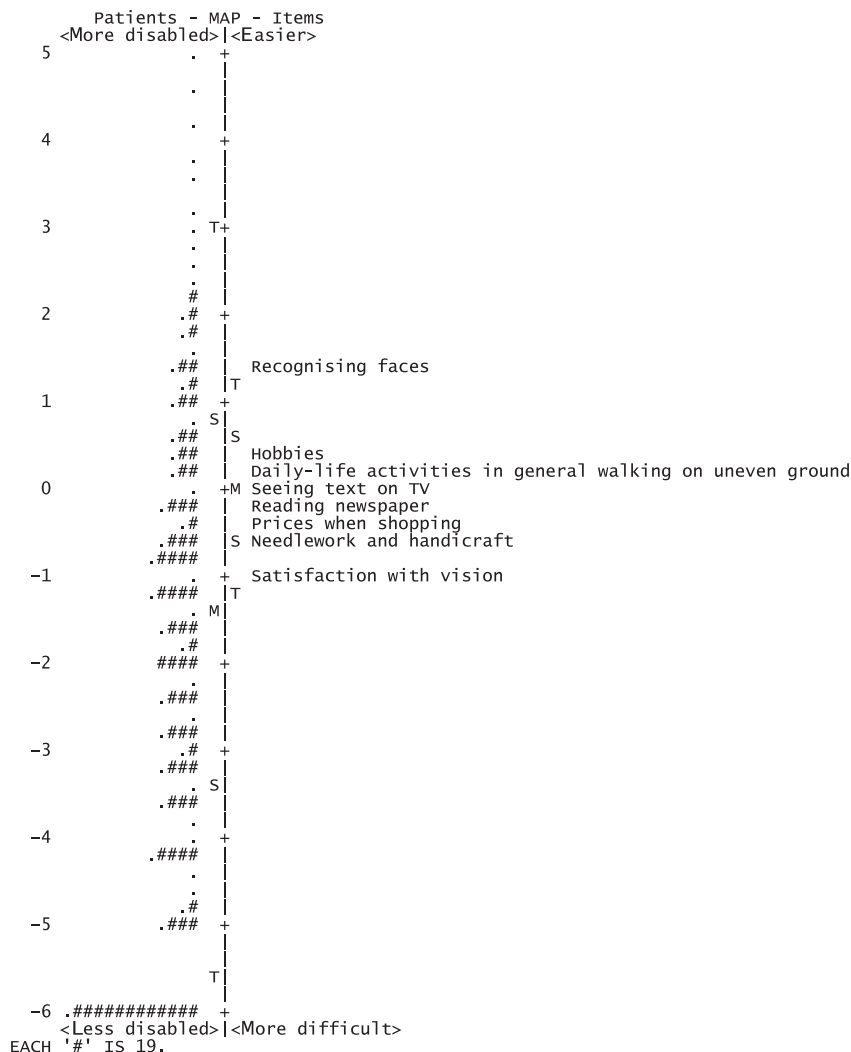
Completed questionnaires before and after surgery were received from 846 patients from six surgical units. The mean age of the patients was 75.6 years ( $\pm 9.2$ ) and 63.7% were female. Table 1 outlines the patient characteristics.

## Results

### Rasch analysis

The category thresholds were ordered. Figures 1–3 show the category probability curves for the items for reading, general difficulty, and satisfaction with vision. The real patient separation was 2.58, and patient separation reliability was 0.87. All items fit a single overall construct (infit range 0.79–1.40; outfit range 0.74–1.40). Table 2 shows the infit and outfit for each item along with the items and response categories for the Catquest-9SF.

Unidimensionality was also tested by principal component analysis of the residuals, which showed that the variance explained by the measures was comparable for empirical calculation (62.3%) and by the model (62.2%). The unexplained variance explained by the first contrast was 1.8 Eigenvalue units (7.4%), which is close to the magnitude seen with random data. The two global assessment items correlated with the first contrast



**Fig. 4.** Person-item map of Catquest-9SF. Distribution of Rasch-calibrated participant scores to the left and item locations to the right. The map contains both preoperative and postoperative data, and so the distribution of patient ability is skewed towards the more capable part. M signifies the mean value of patient ability and item difficulty, respectively. S and T are 1 and 2 standard deviations.

(satisfaction with vision: 0.61; difficulties in general: 0.59). Cronbach's  $\alpha$  was 0.94. The items were sufficiently well targeted to the subjects (mean difference  $-1.95$  logits including both preoperative and postoperative data). This means that the difficulty of the tasks addressed in the items was appropriate for the mixed preoperative and postoperative ability of the patients. The patient-item map given in Fig. 4 shows the relationship between item difficulty and patient ability. The two easiest questions were recognizing faces (1.30) and performing a hobby (0.46). The two most difficult questions were doing needlework and other handicrafts ( $-0.61$ ) and satisfaction with vision ( $-1.05$ ). Three items showed a small DIF by

sex. Seeing well enough to walk on uneven ground (0.68 logits) and seeing well enough to do needlework and other handicrafts (0.75 logits) were both rated by women as easier relative to other tasks. Reading (0.63 logits) was rated as easier by men relative to other tasks. DIF was not evident between the groups divided by the presence or absence of ocular comorbidity or by first- or second-eye surgery. DIF was also tested between preoperative and postoperative data sets. Satisfaction with vision (0.99) was rated easier before surgery, and seeing prices (0.62) was rated easier after surgery.

The Rasch analysis of the Catquest-9SF provided not only the item calibrations found in Table 2, but

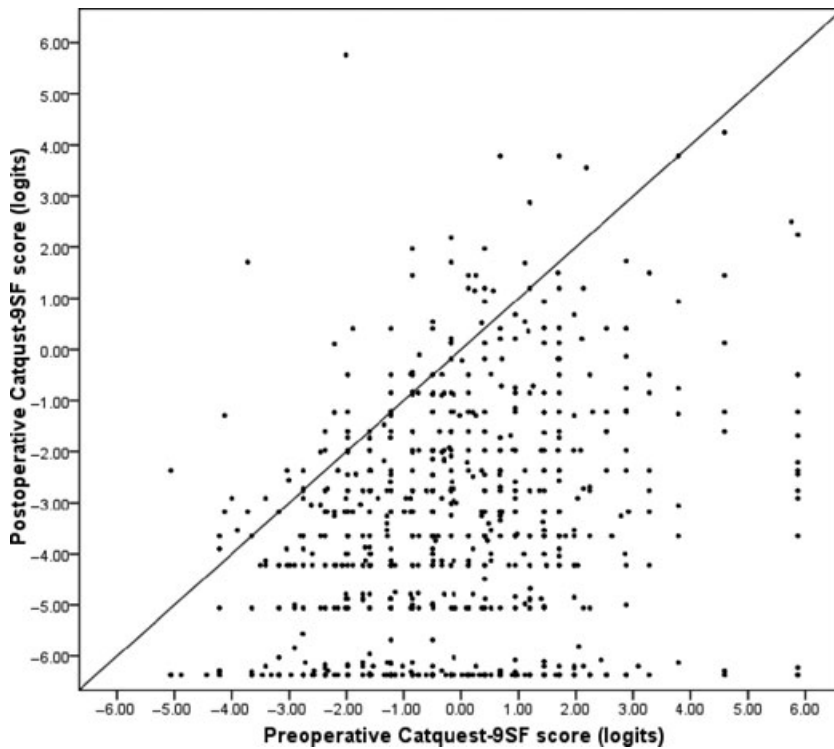
also the item-category calibrations for each of the four response categories of the nine items. These 36 item-category calibrations can be used as anchor values to convert ordinal category values to Rasch measurement estimates. This is valid for both preoperative and postoperative questionnaire data, because the calibrations were developed using a combination of pre 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. A spreadsheet facilitating this conversion is available from the authors.

**Cataract surgery outcomes and validation**

The preoperative Catquest-9SF score correlated with visual acuity in the eye to be operated on ( $r = 0.187$ ), visual acuity in the fellow eye ( $r = 0.208$ ), and visual acuity in the better eye ( $r = 0.224$ ). The postoperative Catquest-9SF score correlated with visual acuity in the operated eye ( $r = 0.361$ ). All four correlations were significant at the 0.01 level (two-tailed). The preoperative mean Catquest-9SF score was  $-0.22 (\pm 1.93)$ , and the postoperative mean Catquest-9SF score was  $-3.69 (\pm 2.28)$ . This 3.47 (95% CI 3.29–3.65) logit improvement after cataract surgery was statistically significant ( $p < 0.0001$ , paired two-tailed  $t$ -test). The change in Catquest-9SF score after cataract surgery means an effect size (improvement/SD) of 1.80. Fig. 5 shows a scatter plot of the preoperative and postoperative Catquest-9SF scores. As is clear from the figure, the majority of patients improved after surgery: 91.5% improved, 1.3% remained in status quo, and 7.2% deteriorated after surgery. The Catquest-9SF scores pre and postoperatively and visual acuity before and after surgery for each participating clinic are shown in Table 3.

**Discussion**

Our findings support the results of the previous revision of the original Catquest and suggestion of a short form – Catquest-9SF (Lundström & Pesudovs 2009). The Rasch analysis of Catquest-9SF shows that the



**Fig. 5.** Scatterplot of preoperative Catquest-9SF score (logits) versus postoperative Catquest-9SF score (logits). A shift in a negative direction between the preoperative and postoperative assessments signifies an improvement.

instrument is a valid measure to be used both before and after cataract extraction. The response categories are ordered, and the instrument has a high precision and is unidimensional as indicated by the fit statistics.

However, some results differ slightly from the original revision and analysis. The original revision was made on data collected between 1995 and 2005 from 58 different surgical units (Lundström & Pesudovs 2009), while the present study concerns data from 2008 involving only six different units. The mean difference between item difficulty and patient ability is larger in this study, meaning that the items, to a certain extent, are deemed easy by many of the patients, at least after sur-

gery. This could be explained by the fact that cataract surgery was performed at an earlier stage of the disease and with less-disabled patients in 2008 compared with 1995–2005. This is supported by the fact that the mean preoperative logMAR (logarithm of the minimum angle of resolution) visual acuity of the surgery eye was 0.46 in the present study but 0.59 in the original study. In addition, mean logMAR visual acuity was better in the present study in the fellow eye preoperatively and the surgery eye postoperatively (0.25 and 0.13 versus 0.27 and 0.15, respectively). However, we do not think that the increased patient ability after surgery causes an issue about the validity of this kind of ques-

tionnaire. The main disadvantage of poor targeting post surgery is that the instrument may underestimate the true benefit of cataract surgery. Targeting could be improved by adding more difficult tasks that better reflect high patient ability; however, this runs the risk of a floor effect pre surgery. If the instrument is supposed to measure the improvement following a surgery that in fact cures the disease, then the items will either be too difficult before surgery or too easy after surgery. In this case, we consider the item difficulty to be reasonably balanced for both preoperative and postoperative measurement. In all other psychometric aspects, the present study and the original one yielded essentially the same results (i.e. item fit, patient separation, unidimensionality, and ordered categories). The Catquest-9SF was reasonably free from DIF in both analyses, and with about similar pattern. The satisfaction item was rated relatively more difficult in the postoperative ranking both in the original study and in this study (1.02 logits and 0.99 logits, respectively).

Six surgical units participated in this study. Four of them were situated in the capital city (Clinics 1, and 4–6) and the other two in the countryside. In the capital city, the surgical rate has been high for many years compared with the rest of the country. Obviously, the patients who underwent surgery in the urban clinics had less self-assessed disabilities than those in the rural clinics (Table 3). The improvement after surgery was much higher in the group that was most disabled preoperatively (clinics 2–3). This is obvious, but it implies that preoperative and postoperative measurements must be interpreted with care. Operating on more disabled cataract patients will give a high improvement score in a measurement like this, but may at the same time

**Table 3.** Visual acuity and Catquest-9SF scores before and after surgery in six surgical units.

Clinic	N	VA surgery eye, preop.	VA better eye, preop.	Catq.-9SF score preop.	VA surgery eye, postop.	Catq.-9SF score postop.	Catq.-9SF score improvement
1	241	0.47 (±0.28)	0.21 (±0.20)	-0.08 (±1.8)	0.18 (±0.24)	-3.35 (±2.38)	-3.27 (±2.66)
2	52	0.49 (±0.24)	0.19 (±0.15)	0.16 (±1.8)	0.11 (±0.12)	-3.91 (±2.06)	-4.08 (±2.41)
3	98	0.45 (±0.25)	0.19 (±0.17)	0.26 (±2.1)	0.10 (±0.16)	-3.99 (±2.16)	-4.26 (±2.64)
4	119	0.43 (±0.25)	0.21 (±0.21)	-0.38 (±2.0)	0.08 (±0.13)	-3.84 (±2.29)	-3.47 (±2.64)
5	194	0.50 (±0.29)	0.18 (±0.19)	-0.58 (±1.9)	0.13 (±0.22)	-3.67 (±2.36)	-3.09 (±2.62)
6	142	0.43 (±0.24)	0.26 (±0.20)	-0.31 (±1.9)	0.13 (±0.19)	-3.86 (±2.08)	-3.55 (±2.47)

VA, Visual acuity, mean logMAR (±SD); Catq.-9SF score = Catquest-9SF score mean value (±SD).

reflect poor availability of surgery and thereby poor health care.

Our results prove that the Catquest-9SF is capable of showing the impact of cataract surgery; in fact, the effect size in this study was 1.80, which is high. The mean preoperative Catquest-9SF score in this study was -0.22 compared with -0.32 in the original study (Lundström & Pesudovs 2009); the corresponding postoperative scores were -3.69 and -3.21, respectively. This means that the average improvement in the present study was a little higher than in the previous study. The percentage of patients with a poorer Catquest-9SF score after surgery was 7.2% in the present study compared with 9.8% in the first study. Thus, there are minor differences between the results in these studies which can be explained by differences in the patient groups. Another possible explanation could be the fact that the follow-up time in the original study was 6 months compared with 3 months in the present study.

The original Catquest questionnaire was developed and validated for cataract patients. The questionnaire has also been used in patients with posterior capsule opacification (Sundelin et al. 2006). The Catquest-9SF could be useful also for patients with corneal or macular diseases, but this has to be proven by validation tests (Rasch analysis).

The small number of items and the restriction to four easily recognized response categories are two of the advantages of Catquest-9SF. Consequently, the burden will be light both on respondents and health care personnel and can easily be incorporated in routine clinical practice. A future development will be to simplify the transfer of data from a paper form or web-based form to a computer program for analysis. Another point to note is that this study of Catquest-9SF was performed with the original Swed-

ish version of the questionnaire, and so the questionnaire text described in this article represents a translation into English. Because of differences in language and culture, further validation tests in other languages are needed if the questionnaire is to be implemented in other countries.

## Acknowledgements

This study was presented in part at the XXVI Congress of the ESCRS, Berlin, Germany, 2008.

## References

Andrich D (1978): A rating scale formulation for ordered response categories. *Psychometrika* **43**: 561–573.

Fisher WP Jr, Eubanks R & Marier RL (1997): Equating the MOS SF36 and the LSU HSI Physical Functioning Scales. *J Outcome Meas* **1**: 329–362.

Garamendi E, Pesudovs K, Stevens MJ & Elliott DB (2006): The Refractive Status and Vision Profile: evaluation of psychometric properties and comparison of Rasch and summated Likert-scaling. *Vision Res* **46**: 1375–1383.

Gothwal VK, Wright T, Lamoureux EL & Pesudovs K (2009): Psychometric properties of visual functioning index using Rasch analysis. *Acta Ophthalmol*. [Epub ahead of print].

Linacre JM (2002): *A User's Guide to Winsteps: Rasch-Model Computer Program*. Chicago: Mesa Press.

Linacre JM (2006): *WINSTEPS Rasch measurement computer program*. Chicago: Winsteps.com.

Lundström M & Pesudovs K (2009): Catquest-9SF patient outcomes questionnaire. Nine-item short-form Rasch-scaled revision of the Catquest questionnaire. *J Cataract Refract Surg* **35**: 504–513.

Lundström M, Jensen S, Fregell G & Roos P (1997): Catquest questionnaire for use in cataract surgery care: description, validity and reliability. *J Cataract Refract Surg* **23**: 1226–1236.

Lundström M, Stenevi U, Thorburn W & Roos P (1998): Catquest questionnaire for use in cataract surgery care: assessment of

surgical outcomes. *J Cataract Refract Surg* **75**: 688–691.

Lundström M, Stenevi U & Thorburn W (2002): The Swedish National Cataract Register: a 9-year review. *Acta Ophthalmol Scand* **80**: 248–257.

Mallinson T, Stelmack J & Velozo C (2004): A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Med Care* **42**: 117–124.

Norquist JM, Fitzpatrick R, Dawson J & Jenkinson C (2004): Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care* **42**: 125–136.

Pesudovs K (2006): Patient-centred measurement in ophthalmology – a paradigm shift. *BMC Ophthalmol* **6**: 25.

Pesudovs K, Burr JM, Harley C & Elliott DB (2007): The development, assessment, and selection of questionnaires. *Optom Vis Sci* **84**: 663–674.

Rasch BG (1960): *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danmarks Paedagogiske Institut.

Sundelin K, Lundström M & Stenevi U (2006): Posterior capsule opacification, a comparison of morphology, visual acuity and self-assessed visual function. *Acta Ophthalmol Scand* **84**: 667–673.

Wolfe EW & Chiu CW (1999): Measuring pretest-posttest change with a Rasch Rating Scale Model. *J Outcome Meas* **3**: 134–161.

Wright BD & Linacre JM (1989): Observations are always ordinal, measurements must be interval. *Arch Phys Med Rehabil* **70**: 857–860.

Received on June 7th, 2009.  
Accepted on October 11th, 2009.

*Correspondence:*  
Mats Lundström  
EyeNet Sweden  
Blekinge Hospital  
SE-371 85 Karlskrona  
Sweden  
Tel: + 46 455 735135  
Fax: + 46 455 20133  
Email: mats.lundstrom@ltblekinge.se