Ophthalmic Epidemiology, 16:296–303
ISSN: 0928-6586 print / 1744-5086 online

DOI: 10.3109/09286580902999454

informa healthcare

Cataract Symptom Score Questionnaire: Rasch Revalidation

Vijaya K. Gothwal,[1,2] Thomas A. Wright,[1] Ecosse L. Lamoureux,[3–5] and **Konrad Pesudovs[1]**

[1] *NH&MRC Centre for Clinical Eye Research, Department of Ophthalmology, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, SA, Australia*
[2] *Meera and L. B. Deshpande Centre for Sight Enhancement, Vision Rehabilitation Centres, L. V. Prasad Eye Institute, Hyderabad, India*
[3] *Centre for Eye Research Australia, Department of Ophthalmology, The University of Melbourne, Victoria, Australia*
[4] *Vision CRC, Sydney, Australia*
[5] *Singapore Eye Research Institute, Singapore National Eye Centre, Singapore*

ABSTRACT

Purpose: **The 5-item Cataract Symptom Score (CSS) was developed using classical test theory to determine appropriate timing for cataract surgery, its outcomes, and whether the symptoms included bother cataract patients. The purpose of this study was to investigate the psychometric properties of the 5-item CSS using Rasch analysis. *Methods:* Responses from 113 patients awaiting cataract surgery to the 5-item CSS (bothered by *double or distorted vision*, *glare or halos*, *blurry vision*, *colors looking different* and *worsening of vision*) were subjected to Rasch analysis. The use of response categories (threshold order), ability of CSS to discriminate between participants' severity of symptoms (person separation, recommended minimum value 2.0), matching of item difficulty to severity of symptoms (targeting), and ability of items to measure a single construct (unidimensionality) were investigated. *Results:* Participants used the response categories as intended. However person separation (1.74) was inadequate, suggesting that the CSS could differentiate only between two groups of participants by their symptoms. Furthermore the CSS was poorly targeted for our population, as the majority of our participants were not bothered by symptoms in the CSS. All items fit the single construct, implying that the CSS is a unidimensional measure of cataract symptoms. *Conclusions:* The CSS is unable to discriminate people with cataract. This problem could be fixed by adding additional questions, but a superior approach may be to create an item bank of cataract symptoms questions, including those of the CSS, and utilize computer-adaptive testing for measurement.**

INTRODUCTION

Over the last few years, patient-reported outcomes (PROs or questionnaires) have become an integral component of the evaluation of the effectiveness of cataract surgery.[1–6] The PROs have thus attained a level of importance comparable with clinical ocular evaluation. Until recently, much of the research using PROs has relied on the traditional theory, the classical test theory (CTT).[1,3,4,6] However the CTT inappropriately treats raw scores and item responses to rating scales (i.e., ordinal data) as interval data.[7–11] Modern psychometric theory, Rasch analysis (based on item response theory, IRT), whereby items and persons can be scaled according to a series of response to items, offers a number of advantages over the CTT. These include identification of unidimensional constructs, additivity of items, and interval-level measurement.[12–14] Unidimensionality means that a single construct is being measured.[15] Additivity refers to the properties of the measurement units (logits), which are the same size (i.e., interval) over the entire continuum if the data fit the model.[14,16] Rasch analysis first converts the questionnaire's ordinal data into interval data, thereby resolving inequalities arising from differential item difficulty in a questionnaire. Furthermore it removes noise from the measurement, which in turn improves sensitivity to change and therefore has advantages for outcomes research.[9,17] In the last few years several researchers have utilized the benefits of Rasch analysis to

Received 14 February 2009; accepted 8 April 2009.
Keywords: Rasch analysis, Cataract, Questionnaire, Cataract Symptom Score
Correspondence to:
Konrad Pesudovs, PhD
NH&MRC Centre for Clinical Eye Research
Department of Ophthalmology, Flinders Medical Centre and Flinders University of South Australia
Bedford Park, South Australia 5042, Australia
E-mail: Konrad.Pesudovs@flinders.edu.au
This manuscript has not been published elsewhere and it has not been submitted simultaneously for publication elsewhere.

either rescale conventionally validated questionnaires using Rasch analysis[18] or reengineer a questionnaire to optimize its performance in ophthalmology.[2,19] However, the Cataract Symptom Score (CSS) is one of the questionnaires that has yet to be validated using Rasch analysis.[3] The CSS is unusual in that very few developed questionnaires assess the symptoms that cause discomfort to cataract patients.[20,21]

Two versions of the CSS exist—one with 5 items[3] and the other with 6 items.[22] Both versions use a four-category Likert design, and the scoring system is based on simple summation of the CSS.[3,20–22] These were developed to help determine appropriate timing for cataract surgery, its outcomes, and whether the symptoms included bothered cataract patients.[20–26] We used the 5-item CSS for the present study, and given the benefits of Rasch analysis, there is justification for subjecting this version of the CSS to the requirements of the Rasch model. The application of Rasch analysis for revalidation has specific advantages for clinical practice. When scale data meet Rasch requirements, the ordinal scores generated from summing item scores can be transformed into interval-level or linear measurements.[2,17,27] These interval measures may be used in a subsequent parametric statistical analyses (for, say, mean scores) that assume an interval-level scale. The aim of the present study was therefore to revalidate the CSS using Rasch analysis.

METHODS

Instrument

Cataract Symptom Score (CSS)

The 5-item CSS consists of five questions (i.e., symptoms that bother cataract patients), and each question has a rating scale of 0 to 4. For each question, a score of 0 was assigned if the patient did not have the symptom or was not at all bothered by it. Scores of 1, 2, and 3 were assigned if the patient was "a little bothered," "somewhat bothered," and "very bothered." The symptoms that comprise the CSS are listed in Table 1.

Subjects

Participants were drawn from the cataract surgery waiting list of the Flinders Medical Centre, Adelaide. Patients were mailed the CSS and the demographic data form (this contained, among other details, the duration of cataract, presence/absence of systemic and ocular comorbidities) prior to their surgery, which they self-administered and returned in a prepaid envelope. The information regarding the comorbidities was then verified from the details filled in by the treating ophthalmologist in the medical records. Included patients were 18 years and older, spoke English without the need for an interpreter, and had no severe cognitive impairment. Typical of a cataract population, patients with coexisting systemic and ocular conditions were included. Ethical approval was obtained, and all patients who agreed to participate signed a consent form. The study was conducted in accordance with the Declaration of Helsinki.

Table 1. Items included in the 5-item Cataract Symptom Score questionnaire

Item no.	Description
1	Bothered by double or distorted vision
2	Bothered by seeing glare, halo, or rings around lights
3	Bothered by blurry vision
4	Bothered by colors looking different from how they used to in a way that is disturbing
5	Bothered by worsening of vision in the last month

Note: The framing question for all items was "Are you . . . ?"

Clinical assessment

Routine clinical assessments were performed prior to cataract extraction. Visual acuity assessments were binocularly performed using computerized testing based on logMAR principles and the screen illumination of 150 cd/m^2. We used habitual binocular acuity, as it is considered to be representative of real-world ability.[28,29] This preoperative visual acuity was recorded approximately 1 month prior to completion of the questionnaire and was extracted from the medical records at the time of data entry.

Rasch analysis

The matrix of responses of 113 patients to the CSS was subjected to Rasch analysis using the Andrich rating scale model for polytomous data.[30] Analyses were performed using the Winsteps software.[31] (version 3.68). Rasch models are a variant of IRT that model a relationship between the level of a latent trait (for the CSS it is severity of symptoms) and the items used for measurement. The concept behind IRT is that participants respond to items in a questionnaire based on the severity of their symptoms (equivalent to person ability in Rasch analysis of a visual disability instrument) and the variation in severity of the symptom questions. Therefore, a person with an average level of severity of symptoms will likely report that they had trouble with the symptoms that are also reported as troublesome by people with greater severity of symptoms and also that they had no trouble with symptoms that are also reported as not troublesome by people with less severity of symptoms. Severity of symptoms is expressed in terms of log odds or "logits," and persons and items are mapped along the same scale. Logit-transformed measures represent linear measures (i.e., the intended severity of the symptoms in this case).[15] For an item, a logit represents the log odds of the severity of an item relative to the severity of the total set of items analyzed. Conventionally, 0 logit is ascribed to the mean item difficulty. For ease of interpretation and consistent with the common notion that logits of higher magnitude represent lower severity of symptoms,[32] we reversed the rating scale for Rasch analysis. Therefore in this study, logits of higher positive magnitude represent a participant who does not have many symptoms due to cataract (and therefore will tend to choose lower response categories such as 0 or 1).

The Rasch model enables examination of the category thresholds. Disordering of thresholds and categories may arise where categories are not as easily understood as intended. Disordered thresholds and categories may affect fit, and consequently categories that are disordered should be collapsed and the Rasch model reapplied to the data. Therefore as a first step in the Rasch analysis we examined the threshold ordering. Next we assessed measurement precision in terms of 'person separation' which gives an estimate of the ability of the questionnaire to discriminate between strata (or groups) of participant's severity of symptoms.[15,33–36] The higher the person separation reliablitity, the more number of strata are differentiated by the questionnaire. A minimum person separation of 2.0 (reliability of 0.8) is considered good that enables the distinction of three strata.[2,15,34]

Rasch analysis also creates an item hierarchy using calibration information based on the correct responses provided by the sample.[15] The item hierarchy can be visualized using the person–item map, and it not only orders items from least to greatest severity of symptoms but also reveals item gaps and redundancies that may lessen the questionnaire's accuracy and efficiency. Comparison of the mean location score obtained for participant's severity of symptoms with that of the value of 0 set for items provides an indication of how well targeted the items are for the participants in the sample. For a well-targeted measure (a balance of easy and difficult items), the mean location for the participants would also be around the value of 0.[37] An absolute mean severity of symptoms of $\geq$0.5 logits indicated mistargeting.

Rasch analysis examines the unidimensionality of the questionnaire, that is, the extent to which all items in the questionnaire measure the same underlying trait (cataract symptoms in this case) and how well each item measures or "fits" the trait.[15] Item fit statistics are an indicator of whether or not each item contributes to the measurement of a single underlying construct.[36,38,39] Item fit to the Rasch model was determined using the mean-square residual fit statistic.[15] The fit statistic used to assess item fit was the infit weighted mean square (MNSQ) statistic that is sensitive to residuals close to the estimated person abilities. It has an expected value of 1.0 and can range from 0 to infinity. Deviations in excess of the expected value may be interpreted as "noise" or lack of fit between the items and the model. Fit values significantly lower than the expected value can be interpreted as item redundancy or overlap. Fit was evaluated against a range of 0.70–1.30 for infit MNSQ.[40] A second strategy to formally test the assumption of unidimensionality involved testing the assumption of local independence by the principal components analysis (PCA) of the residuals.[41] A variance of 60% or greater accounted for by the first factor (i.e., dimension) was considered to be good. The second dimension, or the first contrast in the residuals, can suggest whether there are any patterns in the differences within the residuals large enough to suggest that more than one dimension exists. We used the criterion that the secondary dimension should have the strength of at least two items (as measured by an eigenvalue >2.0) to be considered another dimension, which was greater than the magnitude seen with random data.[42]

Within the framework of Rasch measurement, the questionnaire should work the same way, irrespective of the population subgroup being assessed (e.g., item should behave similarly independent of age, gender, cataract status, etc.). If, at the same severity of symptoms, groups do not display the same probability of endorsing the item, then the item is deemed to display differential item functioning (DIF), thus violating the requirement of unidimensionality.[43–45] We selected the DIF variables a priori for this study. Cataract surgery is performed over a wide age range, and so we included age (<75 years or $\geq$75 years) in the DIF analysis (75 years was the median age of our sample). Gender was included because some items may be gender specific. Cataract surgery may enable participants reduce some symptoms, and so we considered cataract status (i.e., bilateral cataract vs. awaiting surgery in second eye) in our DIF analysis. Participants without systemic or ocular comorbidity may be bothered less by some symptoms and therefore may find it easier to endorse some items than others. So we included systemic and ocular comorbidity in our DIF analyses. DIF analysis can be based either on significance testing or on magnitude. If significance testing is used, the power to detect DIF increases with sample size.[46] We prefer to use magnitude for DIF analysis, and DIF was considered insignificant with the value of <0.50 logits, mild for between 0.50 and 1.00 logits, and notable if >1.00 logits.[47,48] Descriptive statistics were analyzed using SPSS software (version 15.0, SPSS, Chicago, IL).

RESULTS

Of the 195 CSS questionnaires sent, 113 patients completed and returned the CSS questionnaire, resulting in a response rate of 57.9%. The mean age of the patients was 74.8 years (SD = 9.2) and 64 (56.6%) were female. Sixty-four patients (56.6%) were awaiting their first cataract surgery, while 49 (43.4%) were awaiting cataract surgery in their second eye. The demographic characteristics of the participants are provided in Table 2.

Table 2. Sociodemographics of the study population (n = 113)

Characteristic	n (%) or mean ± SD
Age (years)	74.8 ± 9.2
Gender	
Male	49 (43.4%)
Binocular Visual Acuity	
LogMAR, Snellen	0.26 ± 0.21 ($6/12^{+2}$), range −0.26 to 0.92 ($6/3^{-2}$ to $6/48^{-1}$)
Awaiting second-eye surgery	49 (43.4%)
Ocular comorbidity*	
Present	59 (52.2%)
Duration of cataract (years)	
Median (inter-quartile range)	1 (3)
Systemic comorbidity#	
Present	102 (90.3%)

Notes: "*" includes glaucoma, diabetic retinopathy, age-related macular degeneration, etc.; "#" includes diabetes, hypertension, angina, etc.; logMAR = logarithm of minimum angle of resolution; SD = standard deviation.

Table 3. Summary of the Rasch model performance for the person and item parameters for the 5-item CSS

Parameter	Separation	Reliability	Average Infit MNSQ	Model Measurement error	Standard deviation
Persons	1.74	0.75	0.97	0.75	0.15
Items	6.62	0.98	1.01	0.18	0.04

Threshold order and person separation

Category thresholds were ordered. Person separation (and reliability) was poor, indicating that the CSS could discriminate only among two strata (or groups) of participants' severity of symptoms (Table 3).

Unidimensionality and targeting

None of the items did not fit, indicating that all the items of the CSS were measuring a single underlying trait (i.e., severity of symptoms; Table 4). This is reinforced by the PCA of the residuals which showed that the variance explained by the measures was comparable for the empirical calculation (65.3%) and by the model (65.5%). The unexplained variance explained by the first contrast was 1.6 eigenvalue units, which was greater than the magnitude seen with random data. Taken together these findings confirmed the unidimensionality of the CSS. The person–item map (Figure 1) showed a relatively narrow range of item distribution (−2.48 to 0.89 logits) as compared to the participants (range, −5.20 to 4.64 logits) although not significantly different from a normal distribution (Kolmogorov–Smirnov Z test score = 1.06; p = 0.21). The separation between the mean item and participant was >0.50 logits, indicating significant mistargeting. This was further evidenced by a large ceiling effect whereby most of the participants are located in the upper half of the map. The majority of the participants were not typically bothered by the symptoms included in the CSS, and a little over half of participants (58.4%) had no symptoms to discriminate between them, which again illustrates suboptimal targeting of items to participants and is also reflected in the low person separation. Of the five symptoms, only one symptom (*bothered by colors looking different*) was difficult to endorse by all except the most severely symptomatic participants as evidenced by its location below the item mean (mean = 0). Three of the symptoms showed little hierarchical discrimination occurring at approximately the same logit level. The symptom that bothered more of the participants than any other was "*being bothered by blurry vision*" (Figure 1).

Differential item functioning (DIF)

One symptom showed DIF by cataract status. Participants with bilateral cataract rated "*being bothered by worsening of vision*" in the preceding month (0.68 logits) as having lesser impact than the group awaiting cataract surgery in the second eye.

DISCUSSION

The results indicated that the CSS did not possess sufficient discriminatory ability to differentiate between cataract patients as was evidenced by poor person separation. The CSS was only able to differentiate participants into two groups, i.e., lower versus higher severity of symptoms (or less vs. more symptomatic). The associated lower reliability suggested that the user cannot have enough confidence in the item or person estimates. Using the CTT, the CSS was however shown to have high reliability (Cronbach's alpha = 0.82) in a French validation of the CSS.[21] In the CTT, Cronbach's alpha is used as a reliability coefficient to represent the unidimensionality of a questionnaire. According to Cronbach,[49] alpha estimates the "proportion of test variance attributable to common factors among the items," and therefore high inter-item correlations can lead to high Cronbach's alpha.[50,51] Therefore Cronbach's alpha is extremely limited as an indicator of reliability. This limitation highlights the need either to use Rasch analysis in the development stage[17,52,53] or for revalidation of questionnaires in ophthalmology in order to gain a greater insight into questionnaire reliability.[2,18,19,27] The poor person separation in the present study indicates that the 5-item CSS is not very helpful to discriminate between patients (based on their cataract symptoms) on the cataract surgery waiting list in Australia. However, the simplest way to increase person separation would be to add more items to increase the range of symptoms that impact patients with cataract. To address this

Table 4. Item calibration (location) and fit statistics for the five items of the Cataract Symptom Score

Item no.	Item description	Measure (logits)	Standard error (logits)	Infit MNSQ statistic
1	Bothered by double or distorted vision	0.21	0.16	0.90
2	Bothered by seeing glare, halo, or rings around lights	0.73	0.16	1.30
3	Bothered by blurry vision	0.89	0.16	0.72
4	Bothered by colors looking different from how they used to in a way that is disturbing	−2.48	0.25	1.07
5	Bothered by worsening of vision in the last month	0.65	0.16	1.06

Note: All items showed infit mean square values between 0.70 and 1.30.

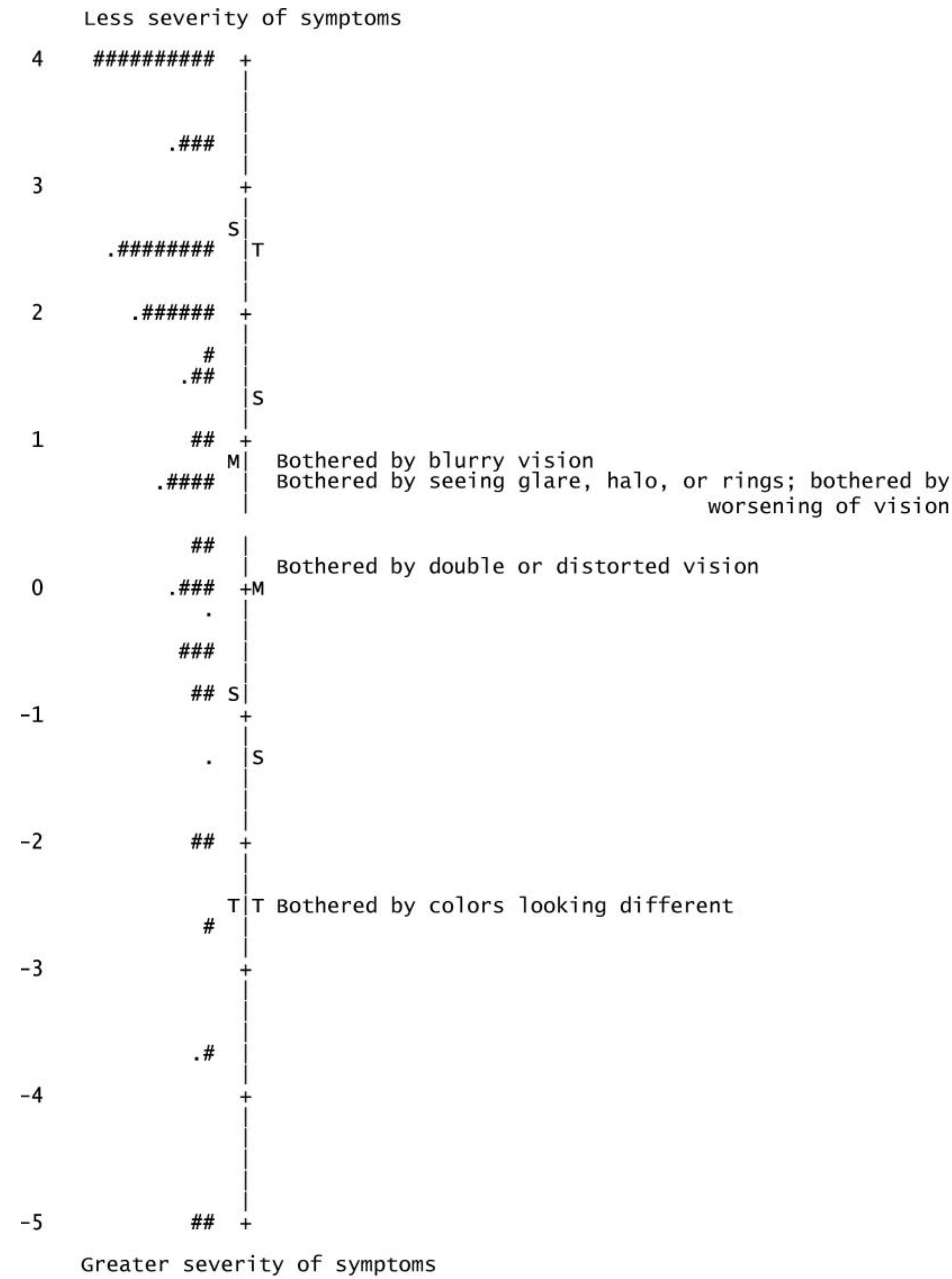

Figure 1. Person-item map for the 5-item Cataract Symptom Score. The participants are located on the left of the dashed line, and the participants with lower severity of symptoms are located at the top of the map. Items (i.e., symptoms) are on the right of the dashed line with the symptoms most easily experienced located toward the top of the map. Each "#" and "." represent two and one participant respectively: M, mean; S, one standard deviation from the mean; T, two standard deviations from the mean.

issue, the use of another version, the 6-item CSS[22] with an extra symptom of "*disturbing brightness*" could have been investigated. However given the relatively large poor separation of the 5-item CSS, it would require addition of at least two more severe symptoms (rather than one extra symptom of the 6-item CSS) to improve its discriminative ability in a cataract population with lower severity of symptoms (i.e., which is less symptomatic) like the sample tested. Care needs to be taken to ensure that the additional symptoms do not overlap with the existing ones and are widely spaced on the severity of symptoms scale. One of the best and recommended ways to find additional symptoms would be to conduct focus group discussions with cataract patients, which should consist of a good mix of patients with varied nature of symptoms so as to capture the entire spectrum of symptoms related to cataract patients.[54] The lack of use of focus groups in the original development of the 5-item CSS may explain the narrow distribution of symptoms.

It is likely that our participants may have experienced bothersome symptoms not addressed by the CSS. For example, the CSS did not include symptoms related to "*seeing faded lane-boundary markings*" or "*glare from oncoming car lights*." Furthermore, the narrow distribution of symptoms limits the usefulness of the questionnaire for our cataract population. Significant gaps in the symptoms were evident, particularly between the items "*bothered by double or distorted vision*" (0.21 logits) and "*being bothered by colors looking different*" (−2.48 logits). Conversely, the other three symptoms were located within approximately 0.15 logits of one another, suggesting redundancy (Figure 1). Furthermore the two items "*bothered by worsening of vision*" and "*bothered by seeing glare, halo or rings around lights*" were located at the same level. Therefore while on one hand there was an irregular and sparse distribution of the five symptoms, on the other hand the existing ones were placed very close to each other.

There was suboptimal targeting indicating that our participants were however not typically bothered by cataract symptoms included in the CSS (Figure 1). This targeting could, however, be expected to further worsen for the CSS after cataract surgery in the second eye, as the cataract symptoms are expected to decrease. Targeting problems have been reported for visual function questionnaires that have been validated using Rasch analysis in a similar cataract population.[2,27,55] Of the symptoms reported by our participants, "*blurry vision*" was the most frequently (66.6% of participants) reported and "*colors looking different*" was the least commonly (78.6% participants reported "*not at all bothered*") reported symptom by our participants. These findings are in line with a study that used the CSS in American and Korean cataract population and found that the former reported no symptoms or not being "bothered" by any symptom significantly than the latter.[20] The difference was attributed to the cultural differences in making decisions on cataract surgery in the two populations. Although matching items to patients' severity of symptoms may miss the opportunity to measure the participants with extreme responses (i.e., those reporting being very bothered), targeting for the CSS is important because when a set of symptoms (or items) is not matched well (i.e., either too hard or easy) to the severity of symptoms, reliability gets affected as was the case.[37]

In conclusion, the Rasch analyses reported here have added another perspective to the 5-item CSS, providing a tentative view of the strengths and weaknesses of the questionnaire. Importantly, the symptoms covered a narrow range of the severity of symptoms, indicating that the 5-item CSS in its present form lacks sufficient ability to discriminate between groups of symptomatic patients, as it does not contain the full spectrum of symptoms that concern cataract patients. We recommend adding at least two more symptoms that would help increase the person separation to satisfactory levels, thereby increasing accurate measurement of the less symptomatic (i.e., with lower severity of symptoms) patients. While this simple modification would significantly improve the psychometric properties of the 5-item CSS, a better option would be to develop new "cataract symptom questionnaire" using Rasch analysis. Comparatively, a superior strategy would be the development of an item bank of symptoms questions,[56] which could be administered by computer-adaptive testing in which questions are asked based on a participant's response to previous items so that the measurement threshold is reached as efficiently as possible. Such item banks exist for other areas of health care,[57–59] and it is time that a similar approach was followed in ophthalmology.

ACKNOWLEDGMENTS

The authors wish to thank all the participants who volunteered to participate in the study.

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

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

1. Mangione CM, Phillips RS, Seddon JM, et al. Development of the "Activities of Daily Vision Scale." A measure of visual functional status. *Med Care.* 1992;30:1111–1126.
2. 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.
3. Steinberg EP, Tielsch JM, Schein OD, et al. The VF-14. An index of functional impairment in patients with cataract. *Arch Ophthalmol.* 1994;112:630–638.
4. Pesudovs K, Coster DJ. An instrument for assessment of subjective visual disability in cataract patients. *Br J Ophthalmol.* 1998;82:617–624.
5. Carta A, Braccio L, Belpoliti M, et al. Self-assessment of the quality of vision: association of questionnaire score with objective clinical tests. *Curr Eye Res.* 1998;17:506–511.

6. Mangione CM, Lee PP, Pitts J, et al. Psychometric properties of the National Eye Institute Visual Function Questionnaire (NEI-VFQ). NEI-VFQ Field Test Investigators. *Arch Ophthalmol.* 1998;116:1496–1504.
7. Nunally JC, Bernstein IH. *Psychometric Theory*, 3rd ed. New York: McGraw-Hill, 1994.
8. Mallinson T. Why measurement matters for measuring patient vision outcomes. *Optom Vis Sci.* 2007;84:675–682.
9. Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C. Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care.* 2004;42:I25–I36.
10. White LJ, Velozo CA. The use of Rasch measurement to improve the Oswestry classification scheme. *Arch Phys Med Rehabil.* 2002;83:822–831.
11. Raczek AE, Ware JE, Bjorner JB, et al. Comparison of Rasch and summated rating scales constructed from SF-36 physical functioning items in seven countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol.* 1998;51:1203–1214.
12. Karabatsos G. The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *J Appl Meas.* 2001;2:389–423.
13. Brogden HE. The Rasch model, the law of comparative judgement and additive conjoint measurement. *Psychometrika.* 1977;42:631–634.
14. Perline R, Wright BD, Wainer H. The Rasch model as additive conjoint measurement. *Appl Psychol Measure.* 1979;3:237–256.
15. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. London: Lawrence Erlbaum, 2001.
16. Fisher GH, Molenaar IW. Rasch models. In: *Foundations, Recent Developments and Applications*. New York: Springer-Verlag, 1995.
17. 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.
18. Lamoureux EL, Pallant JF, Pesudovs K, et al. The Impact of Vision Impairment Questionnaire: An evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci.* 2006;47:4732–4741.
19. 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;4:667–680.
20. Lee JE, Fos PJ, Sung JH, et al. Relationship of cataract symptoms of preoperative patients and vision-related quality of life. *Qual Life Res.* 2005;14:1845–1853.
21. Gresset J, Boisjoly H, Nguyen TQ, et al. Validation of French-language versions of the Visual Functioning Index (VF-14) and the Cataract Symptom Score. *Can J Ophthalmol.* 1997;32:31–37.
22. Schein OD, Steinberg EP, Cassard SD, et al. Predictors of outcome in patients who underwent cataract surgery. *Ophthalmology.* 1995;102:817–823.
23. Brenner MH, Curbow B, Javitt JC, et al. 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.
24. Curbow B, Legro MW, Brenner MH. The influence of patient-related variables in the timing of cataract extraction. *Am J Ophthalmol.* 1993;115:614–622.
25. Javitt JC, Brenner MH, Curbow B, et al. Outcomes of cataract surgery. Improvement in visual acuity and subjective visual function after surgery in the first, second, and both eyes. *Arch Ophthalmol.* 1993;111:686–691.
26. Norregaard JC, Bernth-Petersen P, Andersen TF. Visual impairment and general health among Danish cataract patients. Results from the Danish Cataract Surgery Outcomes Study. I. *Acta Ophthalmol Scand.* 1996;74:598–603.
27. 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.
28. 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.
29. Elliott DB, Hurst MA, Weatherill J. Comparing clinical tests of visual function in cataract with the patient's perceived visual disability. *Eye.* 1990;4(Pt. 5):712–717.
30. Andrich DA. A rating scale formulation for ordered response categories. *Psychometrika.* 1978;43:561–573.
31. Linacre JM. *WINSTEPS Rasch Measurement Computer Program.* Chicago: Winsteps.com, 2008.
32. Ludlow LH, Haley SM. Rasch model logits: Interpretation, use and transformation. *Educ Psychol Meas.* 1995;55:967–975.
33. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989;70:857–860.
34. Duncan PW, Bode RK, Min Lai S, Perera S. Rasch analysis of a new stroke-specific outcome scale: The Stroke Impact Scale. *Arch Phys Med Rehabil.* 2003;84:950–963.
35. Velozo CA, Kielhofner G, Lai JS. The use of Rasch analysis to produce scale-free measurement of functional ability. *Am J Occup Ther.* 1999;53:83–90.
36. Wright BD, Masters GN. *Rating Scale Analysis*. Chicago: MESA Press, 1982.
37. 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–I24.
38. Wright BD, Stone MH. *Best Test Design*. Chicago: MESA Press, 1979.
39. Smith RM. Fit analysis in latent trait measurement models. *J Appl Meas.* 2000;1:199–218.
40. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Meas Trans.* 1994;8:370.
41. Wright BD. Local dependency, correlations and principal components. *Rasch Meas Trans.* 1996;44:46–52.
42. Linacre JM. *User's Guide to Winsteps*. Chicago: MESA Press, 2005.
43. Smith RM, Suh KK. Rasch fit statistics as a test of the invariance of item parameter estimates. *J Appl Meas.* 2003;4:153–163.
44. Scheuneman JD. A method for assessing bias in test items. *J Educ Meas.* 1979;16:143–152.
45. Holland PW, Wainer H. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum, 1993.
46. Wang WC, Yao G, Tsai YJ, et al. Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. *Qual Life Res.* 2006;15:607–620.
47. Wright BD, Douglas GA. *Best Test Design and Self-Tailored Testing*. In: *MESA Research Memorandum No 19*. Chicago: Statistical Laboratory, Department of Education, University of Chicago, 1975.
48. Wright BD, Douglas GA. *Rasch Item Analysis by Hand*. In: *MESA Research Memorandum No 21*. Chicago: Statistical Laboratory, Department of Education, University of Chicago, 1976.
49. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297–334.
50. DeVellis RF. *Scale Development Theory and Application*. Newbury Park, CA: Sage, 1991.
51. Streiner DL, Norman GR. *Health Measurement Scales: A Guide to Their Development and Use*, 3rd ed. New York: Oxford University Press, 2003.
52. 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.

53. Pesudovs K, Garamendi E, Elliott DB. The Contact Lens Impact on Quality of Life (CLIQ) Questionnaire: Development and validation. *Invest Ophthalmol Vis Sci.* 2006;47:2789–2796.

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

55. 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;15:224–233.

56. Wright BD, Bell SR. Item banks: What, why, how. *J Edu Meas.* 1984;21:331–345.

57. Walter OB, Becker J, Bjorner JB, et al. Development and evaluation of a computer adaptive test for "Anxiety" (Anxiety-CAT). *Qual Life Res.* 2007;16(Suppl. 1):143–155.

58. Fliege H, Becker J, Walter OB, et al. Development of a computer-adaptive test for depression (D-CAT). *Qual Life Res.* 2005;14:2277–2291.

59. Lai JS, Cella D, Dineen K, et al. An item bank was created to improve the measurement of cancer-related fatigue. *J Clin Epidemiol.* 2005;58:190–197.