Introduction

Cataract surgery is the most performed ophthalmic surgical procedure worldwide and it is increasingly being offered at lower disease thresholds in Australia, Scandinavia and other developed nations (Keeffe & Taylor 1996; Taylor 2000; Leinonen & Laatikainen 2002). Patients' reported outcomes, including measurement of visual functioning and quality of life, can play an important role in justifying cataract surgery and measuring its outcome (Lundstrom et al. 2001; Mozaffarieh et al. 2005; Lundqvist & Monestam 2008). In recent decades several generic and disease-specific (including cataract) visual function questionnaires (VFQs) have been developed. The majority of these VFQs were developed using classical test theory (CTT), the limitations of which are now acknowledged by most investigators (Massof 2002). The major limitations of CTT relate to the assumptions associated with instrument scoring and the limited ability to assess item characteristics, including whether the items form a unidimensional construct (Hambleton 2000).

The availability of item response theory-based (IRT) models, specifically the Rasch model, has encouraged...
psychometric research to address fundamental measurement issues associated with CTT (McHorney et al. 1997; Hambleton 2000; Tennant et al. 2004). Rasch analysis is the most commonly used IRT model to create new VFQs (Gothwal et al. 2003; Pesudovs et al. 2004; Pesudovs et al. 2006) or test existing VFQs (Velozo et al. 2000; Massof & Fletcher 2001; Pesudovs et al. 2003; Lamoureux et al. 2008). The Rasch measurement models checks two important assumptions: (i) the probability of endorsing one question does not increase the probability of endorsing another one identically (local independence); and (ii) all questions in the questionnaire measure a single underlying construct (unidimensionality). A number of cataract-specific questionnaires [Visual Function-14 (VF-14), Activities of Daily Vision Scale (ADVS), Visual Disability Assessment (VDA)] have been re-examined using Rasch analysis (Velozo et al. 2000; Pesudovs et al. 2003; Pesudovs et al. 2005); the Impact of Cataract Surgery questionnaire (ICS) (Monestam & Wachtmeister 1999) has not.

Therefore, the overall goal of the present study was to use the Rasch model to examine the responses of cataract patients in Australia to the ICS questionnaire.

Materials and Methods

Patients of the Flinders Medical Centre (Adelaide, South Australia) currently on the public waiting list to have cataract surgery participated in this study. Patients were mailed the ICS questionnaire for self-completion, which they returned via a self-addressed envelope. Included patients were aged 18 years or older, English-speaking and had no severe cognitive impairment. Co-existing ocular and systemic comorbidities representative of a typical cataract population in Australia were apparent in the current patient group (Kirkwood et al. 2006). Ethical approval was obtained from the Flinders Clinical Research Ethics Committee and all participants signed a consent form. The study was conducted in accordance with the tenets of the Declaration of Helsinki.

Clinical assessment

Routine clinical assessments occurred prior to listing for cataract extraction. Visual acuity assessments were performed using computerized testing based on LogMAR principles with a screen illumination of 150 cd/m².

The ICS questionnaire

The ICS questionnaire consists of four items (Table 1). The items relate, respectively, to ability to read, watch television, orientate in unfamiliar surroundings and estimate distance (near and far). As can be seen from the table, each item uses a different rating scale.

Rasch analysis

The data were analysed with Winsteps software (Linacre 2008) (Winsteps 3.66; Chicago, Illinois, USA) using the Andrich rating scale model for polytomy data (Andrich 1978). The nature of Rasch analysis has been detailed elsewhere (Massof 2002; Mallinson 2007; Pesudovs et al. 2007). A brief description follows.

The first step was an assessment of the behaviour of the response categories. Because each item had varying number and labelling of the response categories (Table 1), an individual Andrich rating scale was applied for each item format (four) to examine the performance of the response categories. Once the response categories were found to show the intended hierarchy, other characteristics were investigated as follows.

The Winsteps software was used to obtain the estimates of person ability and item difficulties, together with an assessment of measurement precision (using person separation statistics, an indicator of the number of statistically different levels (or strata) of participant ability distinguished by the items; minimum acceptable value of 2.0), unidimensionality [i.e. the extent to which all items in the ICS measure a single underlying construct, reflected in the information-weighted or infit mean square statistics; acceptable fit criterion of 0.7–1.3 and further confirmation of unidimensionality by principal components analysis (PCA) of residuals], differential item functioning (DIF) (i.e. if items perform equally between subgroups; for example, men and women), targeting (the extent to which the difficulty of the items match the abilities of the participants represented by a difference between person and item mean values in the person–item map; a perfectly targeted instrument would have targeting of 0; a difference between means of more than one logit indicates notable mis-targeting) and item hierarchy (i.e. items should form a hierarchy of difficulty, ranging from least to most difficult to perform, also visualized in the person–item map).

The minimum acceptable measurement property for the ICS to be termed as a measure was a person separation $\geq$ 2.0. In case the questionnaire failed this fundamental requirement, further assessments such

Table 1. Contents of the Impact of Cataract Surgery Questionnaire.

<table>
<thead>
<tr>
<th>Item number</th>
<th>Item description</th>
<th>Number of categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Can you read ordinary newspaper-size print? If YES, what visual aids do you need to be able to read?</td>
<td>$4^*$</td>
</tr>
<tr>
<td>2</td>
<td>Do you experience any problems while watching TV caused by your cataractous eye?</td>
<td>$2^*$</td>
</tr>
<tr>
<td>3</td>
<td>Do you experience difficulties when orientating in unfamiliar surroundings?</td>
<td>$3^+$</td>
</tr>
<tr>
<td>4</td>
<td>Do you experience difficulties in estimating distance (nearby/far away)?</td>
<td>$3^+$</td>
</tr>
</tbody>
</table>

* Response options: none or spectacles (1), hand-held or stand magnifiers (2), others (please specify) (3) and no (4, i.e. cannot read).

$^1$ Response options: yes, difficulties (2) and no (1).

$^2$ Response options: no problems (1), some problems (2) and severe problems (3).

$^3$ Response options: No (1), Yes, difficulties to nearby/No to far away (2) or Yes, difficulties to far away/No to near by (2), and Yes, difficulties to nearby and far away (3).
as PCA and DIF were not carried out.

Results

Participant characteristics

Ninety-one participants completed the ICS questionnaire. The mean age of the patients was 74.9 years (range 50–91 years) and 50 (54.9%) were female. Table 2 summarizes the participant characteristics.

Rasch analysis of the ICS data

Assessment of response categories

Figure 1 illustrates the category probability curves (CPCs) of item 1 in its four-point original form showing disordered thresholds necessitating collapsing of categories. The CPC plots visual disability on the x-axis against the probability of endorsing each response category on the y-axis. Threshold refers to the point between two adjacent response categories, for example 1 and 2, where either response (1 or 2) has equal probability of being selected. For a given item, the number of thresholds is always one less than the number of categories. Consequently, one would expect to see three thresholds, but category 3 (others, please specify) was never used and so was not observed. Furthermore, one can see disordered thresholds (i.e. category 2 does not have a range along the scale where it is the most likely category to be selected). Threshold disordered suggests that the response scale is not working adequately to order participants with distinct levels of ability. Therefore, categories were reorganized by collapsing the first two categories to generate dichotomous response categories: 1 represented ‘none/some visual aids (spectacles or hand-held or stand magnifiers)’ and the remaining response option (‘no’) was recoded as 2. The remainder of the item groups demonstrated ordered thresholds.

Person separation and item fit

The person separation for the four-item ICS was unacceptably low (Table 3). Misfit was observed for only one item, ‘Do you experience difficulties in estimating distance nearby/far away?’ (Table 4). However, the person separation did not improve following the deletion of this misfitting item (Table 3).

Item hierarchy and targeting

The person–item map (Fig. 2) illustrates that ‘orientating in unfamiliar surroundings’ was the least difficult activity to perform whereas ‘estimating distance’ was the most difficult. The mean person measure was −0.46 logits (ranging between +3.0 and −3.0 logits), indicating that the item difficulty mostly matched the visual abilities of the participants. In comparison, the distribution of items covered a very narrow range (~1.50 to 1.17 logits). Furthermore, the items were sparsely spread: two items (orientation, reading) were located above the mean item difficulty and the other two (watching TV, estimating distance) were located below the mean item difficulty.

Discussion

Rasch analysis revealed the overall performance of the ICS questionnaire to be poor. The fundamental limitation was inadequate person separation, indicating that the questionnaire was unable to differentiate between the visual abilities of participants awaiting cataract surgery; therefore, it does not supplement the results of a clinical evaluation for cataract surgery. It is ineffective to retain items of no discriminatory value in a questionnaire. However the items of the ICS do constitute important day-to-day activities of elderly patients. Given this, the simplest way to increase the level of discrimination would be to add more items – specifically, those items that cover a wider range of activities such as self-care, face recognition etc., as seen in other VFQs that have good person separation (Velozo et al. 2000; Pesudovs et al. 2003; Pesudovs et al. 2005). Poor person separation can occur for several reasons, but is a common problem with questionnaires containing a smaller number of items. Monestam & Wachtmeister (1999) reported that the ICS was intentionally kept short to ensure simplicity of completion for the elderly population. Although respondent burden is reduced with fewer items (Mallinson et al. 2004), the undue shortening or inclusion of a smaller number of items can disrupt the psychometric properties of the questionnaire and thereby limit its use.

Misfitting item was the other limitation of the ICS. One item (estimating distance) misfit, indicating that this item is not in tandem with the other three items in the measurement of the underlying construct. In other words participants responded to this item erratically, perhaps because the item is

Table 2. Characteristics of respondents to the Impact of Cataract Surgery Questionnaire (n = 91).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n (%) or mean ± standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (year)</td>
<td>74.9 ± 9.0</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>41 (45.1)</td>
</tr>
<tr>
<td>Female</td>
<td>50 (54.9)</td>
</tr>
<tr>
<td>Visual acuity</td>
<td></td>
</tr>
<tr>
<td>Eye to be operated on (habitual)</td>
<td></td>
</tr>
<tr>
<td>LogMAR (Snellen)</td>
<td>0.53 ± 0.35 (6/19–194)</td>
</tr>
<tr>
<td>Range</td>
<td>0.04-1.60 (6/6-7 to 6/240)</td>
</tr>
<tr>
<td>Fellow eye (habitual)</td>
<td></td>
</tr>
<tr>
<td>LogMAR (Snellen)</td>
<td>0.30 ± 0.31 (6/12)</td>
</tr>
<tr>
<td>Range</td>
<td>−0.26 to 1.30 (6/3-2 to 6/12)</td>
</tr>
<tr>
<td>Binocular (habitual)</td>
<td></td>
</tr>
<tr>
<td>LogMAR</td>
<td>0.24 ± 0.21 (6/9.5–1)</td>
</tr>
<tr>
<td>Range</td>
<td>−0.26 to 0.92 (6/3-2 to 6/48–1)</td>
</tr>
<tr>
<td>Awaiting second-eye surgery</td>
<td>40 (44.0)</td>
</tr>
<tr>
<td>Ocular comorbidity†</td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>49 (53.8)</td>
</tr>
<tr>
<td>Absent</td>
<td>42 (46.2)</td>
</tr>
<tr>
<td>Systemic comorbidity‡</td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>10 (11.0)</td>
</tr>
<tr>
<td>Absent</td>
<td>81 (89.0)</td>
</tr>
</tbody>
</table>

† Minus symbol in superscript indicates that a participant missed some letters from that particular line. For example, 6/19 indicates one missed letter from this line.
‡ Includes diabetes, hypertension, angina etc.
not understood well, is ambiguous or measures a second dimension (Pesudovs et al. 2007). Of these, ambiguity appears the most likely in the present case because the item pertains to estimating distances either nearby or far away, which is rather a vague description. Removal of misfitting items usually improves the fit of the model (Pallant et al. 2006). However, this was not the case with the ICS questionnaire and the discriminatory ability continued to remain suboptimal.

The limitations of the performance of VFQs such as the ICS questionnaire could be avoided if Rasch analysis was used in their development, wherein the focus on the 'item' gives a direct connection between the item and the location of the item on the latent variable (Pesudovs et al. 2007). The person–item map of the ICS questionnaire highlights the inadequacies that can be associated with CCT-developed questionnaires, specifically the presence of significant gaps in the locations of the items. This limitation could explain the lack of subjective visual improvement in a proportion of patients (aged 90 years or older) in a cataract-outcomes study that used the modified version of the ICS questionnaire (Monestam & Wachmeister 2004).

Despite these shortcomings, the ICS questionnaire demonstrated one adequate property: targeting. With the exception of the Catquest-9SF (Lundström & Pesudovs 2009), most VFQs that have recently been examined using Rasch analysis have demonstrated poor targeting (Velozo et al. 2000; Pesudovs et al. 2003; Lamoureux et al. 2008; Pesudovs et al. 2008). A visual acuity measurement of 6/15 or 6/12 is considered sufficient for watching television and performing most everyday activities (Bergman & Sjostrand 1992). The mean binocular visual acuity of the participants in the present study was 6/9.5, and the good targeting perhaps suggests that this vision was just sufficient for the participants to perform the activities included in the ICS questionnaire.

Thus, the fundamental limitation of the ICS questionnaire remains its poor discriminatory ability, which could be improved with the addition of appropriately targeted items. Items can be added to static questionnaires; however, an even better strategy would be the creation of item banks for computer adaptive testing (Cook et al. 2005; Fayers 2007; Hays & Lipscomb 2007). Item banks, where items from different questionnaires are pooled, have been created and used in other areas of health assessment (Haley et al. 2004, 2006). It is now time for ophthalmic research to develop such an item bank.
In the meantime, clinicians and researchers can use other VFQs such as the impact of vision impairment (Pesudovs et al. 2008), vision core measure 1 (Lamoureux et al. 2008) or the Catquest-9SF (Lundström & Pesudovs 2009), which have a relatively larger number of items covering a wider range of the construct and have been demonstrated to fulfill the stringent requirements of the Rasch model for the assessment of cataract outcomes.

In conclusion, the ICS questionnaire does not meet the requirements of the Rasch model and thus, in its present form, it appears unsuitable for measuring visual disability in patients awaiting cataract surgery. The fundamental limitation of the ICS questionnaire, its inability to distinguish between the visual abilities of patients with cataract, makes it limited as an outcome measure. Other, better VFQs can be used for the assessment of cataract outcomes.

Acknowledgements

The authors thank all the participants who volunteered to participate in this study. They acknowledge the support of the 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.

References


Received on March 9th, 2009. Accepted on July 27th, 2009.

Correspondence:
Konrad Pesudovs
National Health and Medical Research Council Centre for Clinical Eye Research
Department of Ophthalmology
Flinders Medical Centre
Bedford Park
5042 South Australia
Australia
Tel: + 618 8204 4899
Fax: + 618 8277 0899
Email: konrad.pesudovs@flinders.edu.au