Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 46 (2006) 1375–1383
Vision Research
www.elsevier.com/locate/visres

The Refractive Status and Vision Profile: Evaluation of psychometric properties and comparison of Rasch and summated Likert-scaling ☆

Estibaliz Garamendi [a], Konrad Pesudovs [b], Michael J. Stevens [a], David B. Elliott [a,*]

[a] *Department of Optometry, University of Bradford, Richmond Road, Bradford, West Yorkshire, BD7 1DP, United Kingdom*
[b] *NH and MRC, Centre for Clinical Eye Research, Department of Ophthalmology, Flinders Medical Centre and Flinders University, Bedford Park, SA 5042, Australia*

Received 24 March 2005; received in revised form 5 July 2005

Abstract

The psychometric properties of the Refractive Status and Vision Profile (RSVP) questionnaire were evaluated using Rasch analysis. Ninety-one myopic patients from a refractive surgery clinic and general optometric practice completed the RSVP. Rasch analysis of the RSVP ordinal data was performed to examine for unidimensionality and item reduction. The traditional Likert-scoring system was compared with a Rasch-scored RSVP and a reduced item Rasch-scored RSVP. Rasch analysis of the original RSVP showed poor targeting of item difficulty to patient quality of life, items with a ceiling effect and underutilized response categories. Combining the underutilized response scales and removal of redundant and misfitting items improved the internal consistency and targeting of the RSVP, and the reduced 20-item Rasch scored RSVP showed greater relative precision over standard Likert scoring in discriminating between the two subject groups. A Rasch scaled quality of life questionnaire is recommended for use in refractive outcomes research.

Keywords: Rasch analysis; Quality of life; Validity; Psychometric analysis; Questionnaire

1. Introduction

Refractive error affects over 50% of the UK population (Mintel, 2002). Although spectacles and contact lenses are the primary choice of refractive error correction among myopic patients, during the last decade refractive surgery has gained interest even among successful contact lens wearers (Migneco & Pepose, 1996). Traditionally, the success or failure of refractive surgery has been evaluated by standard clinical measures, such as postoperative uncorrected visual acuity and residual refractive error (Koch, Kohnen, Obstbaum, & Rosen, 1998). However, these measures do not necessarily correlate well with patients' postoperative subjective impressions (Mangione, Lee, & Hays, 1996; McGhee, Craig, Sachdev, Weed, & Brown, 2000; Scott et al., 1994), so the assessment of quality of life (QoL) has become one of the more important outcome measures in refractive surgery and refractive correction clinical research and practice (Fraenkel et al., 2004; Hays et al., 2003; Pesudovs, Garamendi, & Elliott, 2004; Vitale, Schein, Meinert, & Steinberg, 2000).

This increase in attention to quality of life issues related to the correction of refractive error has led to the development and validation of several QoL instruments, such as: the Refractive Status and Vision Profile

☆ Konrad Pesudovs is supported by National Health and Medical Research Council (Canberra, Australia) Sir Neil Hamilton Fairley Fellowship 0061.
* Corresponding author. Tel.:+44 0127 423 4642.
E-mail address: D.Elliott1@bradford.ac.uk (D.B. Elliott).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.07.007

(RSVP) (Vitale et al., 2000), the National Eye Institute Refractive Quality of Life (NEI-RQL) (Berry, Mangione, Lindblad, & McDonnell, 2003) and the Quality of Life Impact of Refractive Correction (QIRC) (Pesudovs et al., 2004). Other studies have reported changes in QoL due to refractive correction, but these results were established with non-validated instruments (Bailey, Mitchell, Dhaliwal, Boxer Wachler, & Zadnik, 2003; Hammond, Puri, & Ambati, 2004; Hill, 2002; McGhee et al., 2000; Rose et al., 2000). The RSVP and NEI-RQL have shown to be sensitive to QoL changes related to visual functioning and refractive error, and have reported improved QoL after refractive surgery (McDonnell et al., 2003; Schein, Vitale, Cassard, & Steinberg, 2001). However, the NEI-RQL, has been shown to be insensitive to differences in QoL between contact lens and spectacle wearers, although it could differentiate both from emmetropes (Nichols, Mitchell, Saracino, & Zadnik, 2003). Similarly, the RSVP has been reported to be insensitive to differences in QoL between different types of contact lenses (Nichols, Mitchell, & Zadnik, 2001) although a larger sample size may have revealed differences (Vitale & Schein, 2002).

Considerations in selecting a QoL instrument should include its reliability and validity. The RSVP and NEI-RQL instruments use traditional Likert scoring (Likert, 1932) where patients' response scores for a selected set of items are summed to derive the overall score. Likert scoring assumes the value of each item represents equal difficulty and it scores them equally. In addition, the linear response scale used for each item assumes uniform changes across categories. For example, in a Likert scaled vision disability instrument such as the Activities of Daily Vision Scale (ADVS) (Mangione et al., 1992), a response of "a little difficulty" (score of 4) is used to represent twice the level of ability as "extreme difficulty" (score of 2) which is similarly two times as good as "unable to perform the activity due to vision" (score of 1) for all items. This appears inappropriate and Rasch analysis has been used to confirm that differently weighted response scales are required for different items to provide a valid scale (Pesudovs, Garamendi, Keeves, & Elliott, 2003). Similarly, Likert scales assume that all items are of equal difficulty. For example, with the ADVS instrument an answer of "a little difficulty" to the question regarding visual difficulties 'driving at night' scores the same as "a little difficulty" with 'driving during the day.' Again, this seems inappropriate and Rasch analysis has been used to confirm that subjects report that 'driving at night' is a more difficult task than 'driving during the day' and Rasch analysis can provide an appropriate weighting for each item (Pesudovs et al., 2003). This new approach to instrument scoring and development using modern psychometric methods, such as Rasch analysis (de Boer, Moll, de Vet, Terwee, & Volker-Dieben, 2004; Fisher, 1994; Fisher, Eubanks, & Marier, 1997; Massof, 2002; Wright & Linacre, 1989) to measure health outcomes has suggested improved validity in item inclusion and on assessment of item difficulty across person QoL (Norquist, Fitzpatrick, Dawson, & Jenkinson, 2004; Pesudovs et al., 2003; Raczek et al., 1998; White & Velozo, 2002).

The purpose of this study was to evaluate the psychometric properties of the RSVP questionnaire using Rasch analysis. We hypothesize that the RSVP, like other non-Rasch developed questionnaires, would have (1) sub-optimal internal consistency and would benefit from item reduction; (2) under performing response categories and would benefit from response scale collapse; and (3) better precision with Rasch scaling. Rasch analysis was used to estimate person and item calibrations on the same scale. It was also used to identify poorly performing questions within the RSVP, which perhaps could be omitted to improve its discriminative ability (Pesudovs et al., 2003). In addition, we compared the discriminative ability of the original Likert-scored RSVP, a Rasch-scored RSVP and a reduced item Rasch-scored RSVP by means of relative precision (Fitzpatrick et al., 2004; McHorney, Haley, & Ware, 1997; Norquist et al., 2004; Raczek et al., 1998). Finally the reanalysed RSVP was compared to the original using the psychometric analyses approach suggested by de Boer et al. (2004).

2. Methods

2.1. Study sample

One hundred and eighty-two patients were randomly recruited from a refractive surgery clinic (RSC, $n = 91$) and general optometric practice (GOP, $n = 91$) in Leeds (UK). The RSVP questionnaire was given to each subject in the RSC and the GOP by their clinicians. RSC patients were requested to complete the questionnaire before they underwent their refractive surgery consultation. Subjects gave informed consent to participate after the nature of the study had been fully described. The study followed the tenets of the Declaration of Helsinki and it was approved by the university's ethics committee. Exclusion criteria were ocular disease, ocular surgery, neurological or systemic disease, any medication that could alter visual function and an inability to read and understand written English. The study was limited to myopic patients with greater than –1.00 D along at least one meridian and a minimum age of 18 years.

Refractive error measure/modality and demographic information were extracted either from respondent's answers on the background information section of the questionnaire or from their clinical records. To determine the socio-economic status of both sample populations, each subject was requested to note their postcode details. A socio-economic status indicator was categorized for each subject based on the 'Indices of Deprivation 2000, Index of Multiple Deprivation specified by the National Statistics website (Office for National Statistics, 2001) and calculated from their postcode. The distribution of the Index of Multiple Deprivation values was skewed, so a logarithmic conversion of these values was performed for comparison of the two samples.

2.2. Instrument

Quality of life was assessed using the RSVP questionnaire. The 42 items of the RSVP (Table 1) are distributed among the domains of concern (6 items), expectations (2 items), physical/social functioning (11 items), driving (3 items), symptoms (5 items), glare (3

Table 1
Assessment of psychometric properties of the RSVP from 182 patients

Items	Skew	Kurtosis	Missing data (%)	Ceiling effect (%)	Infit mean square	Outfit mean square	Item calibration (SE)
20-item scale							
17—I worry about my vision	0.19	−0.23	0.6	11.5	0.89	0.91	39.6 (0.6)
18—My vision is a concern in life	0.09	−0.59	0	7.1	0.97	1.00	36.5 (0.6)
19—My vision holds me back	0.10	−0.82	0	21.4	0.88	0.90	41.1 (0.6)
20—I am frustrated with my vision	−0.02	−0.62	0.6	13.7	0.83	0.84	38.2 (0.6)
21—My vision makes me less self-sufficient	0.53	−0.66	0	35.2	1.02	0.96	44.8 (0.6)
22—Because of my vision, there are things I am afraid to do	0.84	−0.02	0	41.7	1.07	0.99	46.3 (0.7)
23—I could accept less than perfect vision if I didn't need glasses or contact lenses any more	0.17	−0.97	2.7	12.6	1.56	1.71	38.8 (0.6)
24—As long as I could see well enough to drive without glasses or contact lenses, I wouldn't mind having vision that was less than perfect	−0.04	−1.17	1.1	15.4	1.63	1.70	38.6 (0.6)
26—Playing or working outside	1.66	3.01	6.6	56	1.02	0.97	50.8 (0.8)
31—Doing your job	1.84	3.82	6	60.4	0.95	0.83	51.9 (0.9)
32—Doing sports/recreation	0.73	−0.38	13.7	35.2	1.31	1.25	44.1 (0.7)
37—Driving when it is raining	1.31	1.32	11.5	48.4	0.87	0.81	49.3 (0.8)
38—Driving when there is glare from oncoming headlights	0.85	0.32	12.1	33.5	0.86	0.87	45.9 (0.7)
39—Your eyes feeling irritated	0.84	0.38	3.3	36.3	0.82	0.92	46.8 (0.7)
40—Drafts blowing in your eyes	1.10	0.87	11	41.2	0.88	0.83	47.7 (0.8)
45—Glare	1.56	2.41	6.6	56.6	0.96	1.01	51.3 (0.8)
52—Glasses getting dirty or scratched	0.33	−0.63	6	18.7	0.99	1.04	42.3 (0.6)
53—Glasses getting fogged up or wet	0.26	−0.41	7.1	13.2	0.97	1.06	40.8 (0.6)
56—The sensation of having contact lenses in your eyes	1.36	1.59	50.5	27	1.04	0.97	50.4 (1.1)
57—Not being able to wear contact lenses as long as you need to	0.67	−0.49	50	18.7	1.06	1.02	45.4 (0.9)
42-item scale							
25—Watching TV or movies	1.41	1.20	2.7	64.8	0.73	0.69	53.3 (0.9)
27—Taking care or playing with children	1.54	1.54	33	45.6	0.78	0.64	53.0 (1.1)
28—Seeing your alarm clock	2.87	8.35	9.3	75.8	1.14	0.82	58.0 (1.2)
29—Seeing clearly when you wake up	1.80	2.33	18.1	60	1.07	0.96	53.7 (1.0)
30—Seeing a clock on the wall	1.99	3.46	4.4	73.1	0.85	0.67	55.9 (1.1)
33—Swimming	0.10	−1.38	44	14.3	1.97	1.97	37.9 (0.8)
34—Your social life	1.77	3.24	6.6	61	0.98	0.85	51.7 (0.9)
35—Reading and near work	1.36	0.95	7.7	56.6	1.30	1.31	49.7 (0.8)
36—Driving at night	1.15	1.18	11.5	44	0.78	0.73	48.8 (0.8)
41—Eyes being sensitive to light	2.03	4.59	6.6	64.3	0.97	0.99	53.5 (0.9)
42—Pain in your eyes	2.67	8.47	8.8	70.9	1.21	1.11	55.8 (1.1)
43—Changes in your vision during the day	1.89	4.22	6	62.63	0.78	0.70	53.3 (0.9)
44—Your vision being cloudy or foggy	2.10	5.05	6.6	67.6	0.94	0.83	54.6 (1.0)
46—Things looking different out of one eye versus the other	1.94	4.02	9.3	63.2	1.04	0.96	53.4 (1.0)
47—Seeing a halo around lights	1.65	1.96	12.1	62.6	1.00	0.91	54.0 (1.0)
48—Seeing in dim light	1.30	0.94	6.6	58.2	0.78	0.77	52.0 (0.9)
49—Your depth perception	2.26	5.40	10.4	70.3	0.89	0.78	57.1 (1.2)
50—Things appearing distorted	2.56	7.14	6	75.3	0.85	0.73	58.0 (1.2)
51—Judging distance when going up or down steps	2.72	7.57	6.6	76.4	1.20	0.84	57.8 (1.2)
54—Contact lenses popping out/falling out of your eye	2.45	6.33	51.1	36.8	1.33	1.07	55.7 (1.4)
55—Contact lenses getting caught up under your eyelid, or moving around in your eye	1.89	4.11	50.5	29.7	1.19	1.12	52.6 (1.2)
58—Losing a contact lens	2.38	5.16	53.3	36.3	1.59	1.26	55.3 (1.4)

Values of skew, kurtosis, percentage of missing data, percentage ceiling effect [percentage answers in the '*never* (*or equivalent*)' response], Rasch analysis fit statistics (infit and outfit mean square), item calibration and standard error are represented for the 42-item, 5-response scale model. The retained 20-item version and the additional items of the original 42-item RSVP are shown.

items), optical problems (5 items) and problems with corrective lenses (7 items). Responses were scored in accordance with the authors' recommendations (Vitale et al., 2000).

Rating responses of 0 'not applicable' and 6 'never did the activity for other reasons not related to vision (or equivalent)' were scored as missing data. For items related to physical/social functioning, driving, symptoms, glare and optical problems, subjects were requested to answer for when they were wearing spectacles, contact lenses or no correction. Rating scores for patients wearing both spectacles and contact lenses were taken from the worst rating score (Vitale et al., 2000).

2.3. Data analysis

The usefulness of the data provided by the items of the RSVP was analysed using the percentage of missing data, skew and kurtosis (agreement with normality), ceiling effect (percentage of responses on the most able end category of the response scale) and infit/outfit statistics provided by Rasch analysis. Infit mean square (information-weighted fit statistic) and Outfit mean square (outlier-sensitive fit statistic) determine the extent that the observed responses fit the expected responses from the Rasch model. The expected value of both the Infit and Outfit mean square is 1. An infit and outfit mean square value lower than 0.80 is considered as overfitting suggesting the items may be providing redundant information. Mean square values greater than 1.20 are considered as misfitting suggesting the item is not as closely related to the overall construct (QoL in this case) as expected and, therefore, it measures something different (Massof & Fletcher, 2001; Pesudovs et al., 2003; Smith, 1986; White & Velozo, 2002). Responses of the patients to the items of the RSVP were also examined by means of Cronbach's α, Cronbach's *r*-bar and person separation and reliability (Mallinson, Stelmack, & Velozo, 2004).

In addition, Rasch analysis was used to rescale the RSVP and this scoring method was compared to the traditional Likert scoring system. Rasch analysis was performed using Winsteps version 3.35 applying the Andrich rating scale model using joint maximum-likelihood estimation (Andrich, 1978).

The standard scoring system of the RSVP uses Likert scoring which assumes that every item has the same value and the sum of the 42 items scores were calculated (Vitale et al., 2000). To facilitate comparison, both Rasch and Likert scores were linearly transformed to a 0–100 scale, with 0 indicating the best quality of life. To compare the Rasch and Likert scores of the RSVP we determined which scoring method best discriminated between the RSC and GOP groups using the method of relative precision (Norquist et al., 2004; Raczek et al., 1998). We would expect these groups to be different as it has previously been shown using the QIRC questionnaire that patients presenting for refractive surgery feel their QoL is more affected by their refractive correction than patients attending general optometric practice (Garamendi, Pesudovs, & Elliott, in press; Pesudovs et al., 2004). Relative precision coefficients were estimated from the ratio of pairwise *F* statistics (*F* value indicating the discriminative ability of one questionnaire divided by the *F* value indicating the discriminative ability of another questionnaire). Relative precision was calculated for the original Likert-scored RSVP, for a Rasch-scored RSVP and a reduced-item Rasch-scored RSVP that only included the items that were identified as the most useful items by Rasch analysis.

Finally, test–retest reliability of the RSVP was assessed using data from a random sample of 41 patients (19 RSC, 22 GOP; mean test–retest time 3.2 ± 1.0 weeks) using the intraclass correlation coefficient (ICC) (Shrout & Fleiss, 1979) and coefficient of repeatability (COR) (Bland & Altman, 1986).

3. Results

The demographic characteristics of the study population are shown in Table 2. There were no differences between the RSC and GOP groups in terms of age ($p = 0.22$), socio-economic status ($p = 0.74$) and gender ($p = 0.94$). The RSC group had 1.25D more myopia ($p < 0.01$) than the GOP group.

3.1. Descriptive statistics

RSVP data quality is shown in Table 1, which lists skew and kurtosis values (agreement with normality), the percentage of missing data, ceiling effect (percentage of responses on the most able end category of the response scale) and infit/outfit mean square statistics for each item of the RSVP questionnaire.

3.2. Rasch analysis

Fig. 1 shows patient quality of life against item difficulty map established by Rasch analysis for the 42-item RSVP questionnaire. Patient and item scales were linearly transformed from logit units to a 0–100 scale (Umean = 49.179, Uscale = 7.565). Patients (#'s on the left of the map) with better quality of life and more difficult items (labelled on the right) appear in descending order from the top to the bottom of the map. Items, on the right, are represented as item numbers with a decimal representing the response scale boundary. With a 5-category scale there are four boundaries between categories so that each item is represented in the figure by four points, being the point on the scale where the response most likely to be selected changes from one

Table 2
Demographic characteristics of the 182 subjects from both refractive surgery clinic ($n = 91$) and general optometric practice ($n = 91$) sample groups

Characteristics	Refractive surgery clinic	General optometric practice
Age (mean ± SD; range), years	39.2 ± 10.7 (20–64)	37.2 ± 10.8 (18–67)
Gender (% female)	58	62
Socio-economic status	1.34 ± 0.25	1.35 ± 0.23
Race (% white)	100	100
Spectacle wearers (%)	57	48
Mean sphere correction (median DS, range)	−4.25 (−1.25 to −18.50)	−3.00 (−0.25 to −11.00)

Socio-economic status was determined by using the 'Indices of Deprivation 2000, Index of Multiple Deprivation' specified by the National Statistics website (Office for National Statistics, 2001).

```
        poorer QoL   less difficult items
                   |
80                 +
                   |
                   |
                   |  28.5 50.5
                   |  49.5 51.5
70                 +  30.5 42.5 44.5 47.5 54.5 58.5
                   |  25.5 27.5 29.5 41.5 43.5 46.5 48.5 55.5
                   |  26.5 31.5 34.5 45.5 56.5
                .  |  35.5 36.5 37.5
                   |T 28.4 22.5 50.4 39.5 51.4 40.5
60                 +  30.4 21.5 42.4 32.5 49.4 38.5 54.4 57.5
                   |  29.4 52.5 44.4 47.4 58.4
                   |S 25.4 19.5 27.4 53.5 31.4 34.4 41.4 43.4 46.4 48.4 55.4
                   |  28.3 26.4 17.5 50.3 35.4 20.5 51.3 45.4 23.5 56.4 24.5
                .  |  30.3 36.4 18.5 42.3 37.4 33.5 49.3 40.4 54.3
50                 +M 29.3 22.4 41.3 38.4 44.3 39.4 46.3 47.3 58.3
                # T|  25.3 21.4 27.3 32.4 31.3 57.4 34.3 43.3 48.3 55.3
               .#  |  26.3 52.4 35.3 45.3 56.3
              ###  |  28.2 36.3 17.4 50.2 37.3 19.4 51.2 40.3 53.4
             .### S|S 30.2 22.3 20.4 42.2 38.3 23.4 49.2 39.3 24.4 54.2 57.3 33.4
40        .######  +  25.2 21.3 18.4 29.2 32.3 41.2 43.2 44.2 46.2 47.2  58.2
        #########  |  27.2 52.3 31.2 34.2 45.2 48.2 55.2
    ############# M|T 26.2 17.3 35.2 19.3 56.2 53.3
        .########  |  36.2 20.3 37.2 23.3 40.2 24.3 33.3
            .####  |  22.2 18.3 38.2 39.2 57.2
30           .### S+  21.2 32.2
             .###  |  52.2
                . T|  17.2 19.2 53.2
                .  |  20.2 23.2 24.2 33.2
                .  |  18.2
20              .  +
                   |
                   |
                   |
                   |
10                 +
                   |
                   |
                   |
                   |
 0                 +
        better QoL   more difficult items

        Each  '#' is 3 patients
```

Fig. 1. Patient QoL/item difficulty map for the 42-item RSVP. On the left of the dashed line are the patients, represented by #. On the right are the crossover points between each response scale (level of the scale where the answer category is most probable to be rated by a patient with that QoL). Patients with better quality of life and more difficult items are near the bottom of the map; patients with poorer quality of life and less difficult items are near the top. The scale is in units 0–100. (M) mean; (S) 1SD from the mean; (T) 2SD from the mean.

category to the next. For this sample, the items, placed higher, are too easy for the patients which are placed lower on the map. This suggests that item difficulty poorly targets patient quality of life issues related to refractive error correction. The patient and item means, indicated by M in Fig. 1, are separated by 12.86 units quantifying poor item to person targeting.

3.3. Response scale reduction

The 5-rating categories on the RSVP were not utilized with the same frequency across all 182 patients. The "always" or equivalent response category (5 response category) was not used for 10 of the 42 items by any of the patients. Responses on this end category differed from 0.5% to 22.5% of which only one item had 22.5% of the responses and all the remaining items had lower than 10% of the responses. To improve the underutilization of the end category the two highest rating categories "always" (5) and "often" (4) could be combined (Linacre, 1999; Pesudovs et al., 2004). This resulted in a reduction in the mean difference between item difficulty and patient quality of life from 12.86 to 9.82 units and a slight increase in patient separation from 2.71 to 2.76.

Person separation is an indicator of the ability of the instrument to differentiate between different patients' quality of life. It is expressed as the ratio of the adjusted standard deviation to the root mean square error (Pesudovs et al., 2003). A person separation value of greater than 2.0 is indicative that patients are significantly different in quality of life across the measurement distribution.

3.4. Item reduction

Although the combined response category improved the difference between item and patient mean values there were still several items providing relatively little information as shown by the data in Table 1. Therefore, we attempted to remove items from the questionnaire that contributed little to the assessment of patient's quality of life. The criteria used for item removal were: (1) infit mean square outside the range 0.80–1.20; (2) outfit mean square outside the range 0.70–1.30; (3) high proportion of missing data; (4) ceiling effect; (5) skew and kurtosis outside the range −2.00 to +2.00 (Pesudovs et al., 2003). Items were removed one at a time as item removal changes fit statistics. This improved the fit of some items that initially had high infit/outfit values and reduced the mean difference between item difficulty and patient quality of life. If removal of an item with high or low infit/outfit values considerably decreased person separation (<2.0) that item was retained. This iterative process finally resulted in a 20-item questionnaire with a person separation of 2.01 and mean difference of 5.16 units (Fig. 2). Items 23 and 24 still had high infit and outfit values and item 20 had low fit values. However, removal of these items decreased person separation to below 2.0 and increased mean difference suggesting poor targeting. Therefore, these three items were retained. This indicates that while the reduced version of the RSVP is improved, it is still not ideal.

Cronbach's α for the original RSVP validation was 0.92 (range from 0.70 to 0.93, Vitale et al., 2000) and for this data set it was 0.98. For the shortened 20-item version of the RSVP questionnaire Cronbach's α was 0.90 (Table 3). Cronbach's *r*-bar, which accounts for the effect of the number of items on Cronbach's α (Massof, 2004) is also provided in Table 3 and shows very low inter-item consistency for the Likert scored RSVP. Test–retest repeatabilities of data from 41 subjects using the same scoring rules are also shown in Table 3.

3.5. Relative precision

The ability of the RSVP to discriminate between subjects considering refractive surgery and those from optometric practice is shown for various Likert and Rasch scoring rules in Table 3, using assessments of relative precision (Norquist et al., 2004; Raczek et al., 1998).

```
                 poorer QoL   less difficult items

                          .   |
 80                           +
                              |
                              |
                              |  31.4
                              |  26.4 45.4 56.4
 70                           +  37.4
                              |  40.4
                              |  22.4 38.4 39.4
                              |T 21.4 32.4 57.4
                          .   |
 60                       # T+   31.3 52.4 45.3
                          .   |  26.3 19.4 56.3 53.4
                        .##   |S 37.3 17.4 20.4 23.4 24.4
                      .####   |  39.3 40.3
                  .####### S|    22.3 18.4 38.3 57.3
 50                 .######   +M 21.3 32.3
               .###########   |  52.3
              .############ M|   26.2 19.3 31.2 53.3 45.2
                .##########   |S 37.2 17.3 56.2
                .##########   |  20.3 23.3 24.3
 40                 #######   +  22.2 18.3 39.2 40.2
                      ##### S|T  21.2 38.2 57.2
                      .####   |  32.2
                         .#   |  52.2
                         ##   |  19.2 53.2
 30                       # T+   17.2
                              |  20.2 23.2 24.2
                          #   |  18.2
                              |
                          .   |
 20                           +
                              |
                              |
                              |
                              |
 10                           +
                              |
                              |
                              |
                              |
  0                           +
                 better QoL   more difficult items

                   Each '#' is 2 patients
```

Fig. 2. Patient QoL/item difficulty map for the 20-item RSVP.

As an example of the calculation of relative precision, Likert scoring indicated mean values (±1 SE) for the RSC and GOP groups of 18.02 ± 1.03 and 23.38 ± 1.22 respectively. ANOVA indicated that these values were significantly different ($F = 11.37$, $p < 0.01$). Rasch 5-category scoring gave mean values of 36.46 ± 0.39 and 38.61 ± 0.47 ($F = 12.46$, $p < 0.01$). Relative precision coefficients were estimated from the ratio of pairwise F statistics. Therefore, relative precision of the Rasch-scored 5-category RSVP was calculated as: relative precision = 12.46/11.37 = 1.10. The 20-item Rasch scaled RSVP had much greater relative precision (2.97). These analyses were repeated using ANCOVA with refractive error as a covariate, given the significant difference in refractive error between the RSC and GOP groups. In this case, the difference in the RSVP Likert scores between the two groups was not significant ($F = 2.74$, $p = 0.10$), whereas it remained highly significant for the Rasch scored 20-item RSVP ($F = 11.25$, $p = 0.01$). Using these ANCOVA F-values provides a relative precision value of 4.11 (11.25/2.74) for the 20-item Rasch scaled RSVP.

Table 3
Person separation and reliability, targeting of items to patients and Cronbach's α and r-bar for the original RSVP validation and the three Rasch scaled versions

	Likert scored RSVP	Rasch scored RSVP (5 responses)	Rasch scored RSVP (4 responses)	Rasch scored 20-item RSVP
Person separation	—	2.71	2.76	2.01
Reliability		0.88	0.88	0.80
Adjusted SD (AdjSD)		5.29	6.19	6.45
Root mean square error (RMSE)		1.95	2.24	3.21
Difference between item and patient means (patient SEM)	—	12.86 (1.95)	9.82 (2.24)	5.16 (3.21)
Cronbach's α	0.92	0.98	0.98	0.90
Cronbach's r-bar	0.22	0.54	0.54	0.47
Relative precision	1.00	1.10	1.18	2.97
Coefficient of repeatability	±12.5	±4.9	±5.4	±5.8
Intraclass correlation coefficient	0.74	0.74	0.74	0.80

4. Discussion

As hypothesised, Rasch analysis illustrated the inadequacy of summated Likert scoring and using traditional methods in the development and validation of questionnaires. Comparison of item difficulty to patient quality of life (Fig. 1) showed poor targeting of item difficulty to patient QoL measures which is reflected numerically by the high mean difference between item difficulty and patient QoL mean values. Rasch analysis also showed that several items provided less useful data as indicated by poor fit statistics. Combining the two highest response categories "always" (5) and "often" (4) or equivalent improved the targeting of item difficulty to patient QoL. While this decreased the mean difference between item difficulty and patient quality of life and improved patient separation, there were still misfitting items that did not discriminate between patients. Reduction of redundant and misfitting items resulted in a 20-item RSVP questionnaire with a person separation over 2.0 and improved targeting of items to patients (Fig. 2). The 20-item version was still not ideal as it had three misfitting items (Infit > 1.20 or <0.80 and Outfit > 1.30 or <0.70) but removal of these items decreased person separation to values under 2.0, which is an unacceptable loss of precision. Removal of these items and addition of other items relevant to patients' quality of life could further improve discrimination and validity. The reduction in respondent burden was estimated as reducing time taken to complete the questionnaire by over half, based on a reduction in the number of questions from 42 to 20.

The original RSVP questionnaire contains 14 items related to physical and social functioning and driving and eight related to symptoms and glare. After using Rasch analysis to remove the least useful items, the reduced item RSVP only included 20-items of which five were related to physical and social functioning and driving and three to symptoms and glare. These results are consistent with our previous report that introduced a refractive correction-related QoL questionnaire, the QIRC questionnaire, in which Rasch analysis identified that patients with corrected refractive error felt they had few problems with visual function and few symptoms, and that issues such as convenience, cost, health concerns and appearance determine the influence of refractive correction on QoL (Pesudovs et al., 2004). Perhaps the reason why the original RSVP was so heavily weighted with functioning and symptoms questions was because the items were principally determined by clinicians (Schein, 2000), who tend to deal with patients' presenting complaints of symptoms or functional difficulties, instead of using an objective approach such as Rasch analysis. The 20-item RSVP still includes a moderate amount of items on the physical and social functioning, driving, symptoms and glare domains but this could be because there is lack of items to discriminate between the remaining domains.

The Likert scored RSVP data had an extremely high value of Cronbach's α of 0.98. Cronbach's α is generally used as a reliability coefficient, to represent the unidimensionality of a questionnaire, but it is dependent on the magnitude of correlations between items and the number of items in the questionnaire (Streiner & Norman, 2003). Therefore inclusion of many highly correlated items provide high values of Cronbach's α (>0.90), which suggest a high degree of item redundancy (DeVellis, 1991; Streiner & Norman, 2003). In addition, Cortina (1993) reported that Cronbach's α is dependent upon the number of items, so that an instrument with more than 20 items would have a high Cronbach's α implying high correlation among items without necessarily having high inter-item correlation at all. Rasch analysis identified redundancy on the original RSVP and removal of more than half of the items lowered Cronbach's α (0.90) to a more acceptable value (Streiner & Norman, 2003) although this is probably still too high. Cronbach's r-bar values (Table 3), which account for the effect of the number of items on Cronbach's α are low (Massof, 2004). In addition, person separation decreased to 2.01 compared to the original Rasch scored RSVP (Table 3) with increased variability in patients'

QoL (AdjSD) and the reduced number of items and rating scale categories (RMSE) (Mallinson et al., 2004).

The relative precision validity coefficient was increased when using the Rasch scoring method over the standard Likert scoring in discriminating between patients from both sample populations. The relative precision was further increased using the 4-category Rasch-scored RSVP and particularly with the 20-item Rasch-scored RSVP. This is likely due to reduced standard error values (Raczek et al., 1998) due to the elimination of poorly performing items and the appropriate weighting of items and categories within items by Rasch analysis and the ability of Rasch analysis to estimate scores despite missing data (Raczek et al., 1998).

The coefficients of repeatability show a much improved repeatability for the Rasch-based scoring methods, so that a significant change in score from test to retest is more than halved when Rasch scoring is used. The lack of change in the test–retest correlation coefficients is due to the influence of the range of scores on correlation coefficients (Bland & Altman, 1986), which is greater for Likert scoring and offsets the poorer repeatability compared to Rasch scoring.

In summary, Rasch analysis was used to assess the psychometric properties of the RSVP questionnaire. As hypothesised, reduction of redundant and misfitting items could improve the internal consistency and targeting of the RSVP, resulting in a 20-item RSVP. Relative precision demonstrated that Rasch-based scoring methods provided a more precise estimate of RSVP scores and improved discrimination between groups. Therefore, we would recommend the use of a Rasch scaled QoL questionnaire for use in refractive outcomes research. Either a questionnaire developed using Rasch analysis such as QIRC, or using the shortened 20-item Rasch-scaled RSVP questionnaire.

Acknowledgment

The authors thank Boots Opticians and Ultralase for providing access to patients.

References

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

Bailey, M. D., Mitchell, G. L., Dhaliwal, D. K., Boxer Wachler, B. S., & Zadnik, K. (2003). Patient satisfaction and visual symptoms after laser in situ keratomileusis. *Ophthalmology, 110*, 1371–1378.

Berry, S., Mangione, C. M., Lindblad, A. S., & McDonnell, P. J. (2003). Development of the National Eye Institute refractive error correction quality of life questionnaire: Focus groups. *Ophthalmology, 110*, 2285–2291.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*, 307–310.

de Boer, M. R., Moll, A. C., de Vet, H. C., Terwee, C. B., Volker-Dieben, H. J., et al. (2004). Psychometric properties of vision-related quality of life questionnaires: A systematic review. *Ophthalmic Physiological Optics, 24*, 257–273.

Cortina, J. M. (1993). What is coefficient alpha?—an examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104.

DeVellis, R. F. (1991). *Scale development theory and applications*. Newbury Park: Sage.

Fisher, W. P. Jr., (1994). The Rasch debate: Validity and revolution in educational measurement. In M. Wilson (Ed.), *Objective measurement: Theory into practice* (pp. 36–72). Norwood NJ: Ablex.

Fisher, W. P., Jr., Eubanks, R. L., & Marier, R. L. (1997). Equating the MOS SF36 and the LSU HSI Physical Functioning Scales. *Journal of Outcome Measurement, 1*, 329–362.

Fitzpatrick, R., Norquist, J. M., Jenkinson, C., Reeves, B. C., Morris, R. W., Murray, D. W., et al. (2004). A comparison of Rasch with Likert scoring to discriminate between patients' evaluations of total hip replacement surgery. *Quality of Life Research, 13*, 331–338.

Fraenkel, G., Comaish, I. F., Lawless, M. A., Kelly, M. R., Dunn, S. M., Byth, K., et al. (2004). Development of a questionnaire to assess subjective vision score in myopes seeking refractive surgery. *Journal of Refractive Surgery, 20*, 10–19.

Garamendi, E., Pesudovs, K., & Elliott, D. B. (2005). Changes in quality of life after LASIK for myopia. *Journal of Cataract and Refractive Surgery*.

Hammond, S. D., Jr, Puri, A. K., & Ambati, B. K. (2004). Quality of vision and patient satisfaction after LASIK. *Current Opinion in Ophthalmology, 15*, 328–332.

Hays, R. D., Mangione, C. M., Ellwein, L., Lindblad, A. S., Spritzer, K. L., & McDonnell, P. J. (2003). Psychometric properties of the National Eye Institute-Refractive Error Quality of Life instrument. *Ophthalmology, 110*, 2292–2301.

Hill, J. C. (2002). An informal satisfaction survey of200 patients after laser in situ keratomileusis. *Journal of Refractive Surgery, 18*, 454–459.

Koch, D. D., Kohnen, T., Obstbaum, S. A., & Rosen, E. S. (1998). Format for reporting refractive surgical data. *Journal of Cataract and Refractive Surgery, 24*, 285–287.

Likert, R. A. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1–55.

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3*, 103–122.

Mallinson, T., Stelmack, J., & Velozo, C. (2004). A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Medical Care, 42*, I17–I24.

Mangione, C. M., Phillips, R. S., Seddon, J. M., Lawrence, M. G., Cook, E. F., Dailey, R., et al. (1992). Development of the 'Activities of Daily Vision Scale'. A measure of visual functional status. *Medical Care, 30*, 1111–1126.

Mangione, C. M., Lee, P. P., & Hays, R. (1996). Measurement of visual functioning and health-related quality of life in eye disease and cataract surgery. In B. Spilker (Ed.), *Quality of Life and Pharmacoeconomics in Clinical Trials* (pp. 1045–1051). Philadelphia: Lippincott–Raven.

Massof, R. W., & Fletcher, D. C. (2001). Evaluation of the NEI visual functioning questionnaire as an interval measure of visual ability in low vision. *Vision Research, 41*, 397–413.

Massof, R. W. (2002). The measurement of vision disability. *Optometry and Vision Science, 79*, 516–552.

Massof, R. W. (2004). Likert and Guttman scaling of visual function rating scale questionnaires. *Ophthalmic Epidemiology, 11*, 381–399.

McDonnell, P. J., Mangione, C., Lee, P., Lindblad, A. S., Spritzer, K. L., Berry, S., et al. (2003). Responsiveness of the National Eye Institute Refractive Error Quality of Life instrument to surgical correction of refractive error. *Ophthalmology, 110*, 2302–2309.

McGhee, C. N., Craig, J. P., Sachdev, N., Weed, K. H., & Brown, A. D. (2000). Functional, psychological, and satisfaction outcomes of laser in situ keratomileusis for high myopia. *Journal of Cataract and Refractive Surgery, 26*, 497–509.

McHorney, C. A., Haley, S. M., & Ware, J. E. Jr., (1997). Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. *Journal of Clinical Epidemiology, 50*, 451–461.

Migneco, M. K., & Pepose, J. S. (1996). Attitudes of successful contact lens wearers toward refractive surgery. *Journal of Refractive Surgery, 12*, 128–133.

Mintel (2002). *Optical goods and eyecare*. London: Mintel.

Nichols, J. J., Mitchell, G. L., & Zadnik, K. (2001). The performance of the refractive status and vision profile survey in a contact lens clinical trial. *Ophthalmology, 108*, 1160–1166.

Nichols, J. J., Mitchell, G. L., Saracino, M., & Zadnik, K. (2003). Reliability and validity of refractive error-specific quality-of-life instruments. *Archives of Ophthalmology, 121*, 1289–1296.

Norquist, J. M., Fitzpatrick, R., Dawson, J., & Jenkinson, C. (2004). Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Medical Care, 42*, 25–36.

Office for National Statistics. (2001). National Statistics Online [National Statistics Website]. 2001. <http://www.statistics.gov.uk/default.asp.> Accessed March 21st.

Pesudovs, K., Garamendi, E., Keeves, J. P., & Elliott, D. B. (2003). The Activities of Daily Vision Scale (ADVS) for cataract surgery outcomes: Re-evaluating validity with Rasch analysis. *Investigative Ophthalmology and Visual Science, 44*, 2892–2899.

Pesudovs, K., Garamendi, E., & Elliott, D. B. (2004). The Quality of Life Impact of Refractive Correction (QIRC) questionnaire: Development and validation. *Optometry and Vision Science, 81*, 769–777.

Raczek, A. E., Ware, J. E., Bjorner, J. B., Gandek, B., Haley, S. M., Aaronson, N. K., et al. (1998). 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. *Journal of Clinical Epidemiology, 51*, 1203–1214.

Rose, K., Harper, R., Tromans, C., Waterman, C., Goldberg, D., Haggerty, C., et al. (2000). Quality of life in myopia. *British Journal of Ophthalmology, 84*, 1031–1034.

Schein, O. D. (2000). The measurement of patient-reported outcomes of refractive surgery: The refractive status and vision profile. *Transactions of the American Ophthalmological Society, 98*, 439–469.

Schein, O. D., Vitale, S., Cassard, S. D., & Steinberg, E. P. (2001). Patient outcomes of refractive surgery: The refractive status and vision profile. *Journal of Cataract and Refractive Surgery, 27*, 665–673.

Scott, I. U., Schein, O. D., West, S., Bandeen-Roche, K., Enger, C., & Folstein, M. F. (1994). Functional status and quality of life measurement among ophthalmic patients. *Archives of Ophthalmology, 112*, 329–335.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Smith, R. M. (1986). Person fit in the Rasch model. *Educational and Psychological Measurement, 46*, 359–372.

Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use* (3rd ed.). New York: Oxford University Press Inc.

Vitale, S., Schein, O. D., Meinert, C. L., & Steinberg, E. P. (2000). The Refractive Status and Vision Profile: A questionnaire to measure vision-related quality of life in person with refractive error. *Ophthalmology, 109*, 1529–1539.

Vitale, S., & Schein, O. D. (2002). Refractive Status and Vision Profile survey in a contact lens trial letter. *Ophthalmology, 107*, 2188–2190.

White, L. J., & Velozo, C. A. (2002). The use of Rasch measurement to improve the Oswestry classification scheme. *Archives of Physiological and Medical Rehabilitation, 83*, 125–136.

Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal, measurements must be interval. *Archives of Physiological and Medical Rehabilitation, 70*, 857–860.