


Validation and comparison of the National Eye Institute Visual Functioning Questionnaire-25 (NEI VFQ-25) and the Visual Function Index-14 (VF-14) in patients with cataracts: a multicentre study

Yu Wan,^{1,†} Liming Zhao,^{2,†} Chen Huang,^{1,†} Yanhui Xu,^{3,†} Min Sun,^{4,†} Yang Yang,^{5,†} Li An,^{6,†} Shuxuan Lv,^{7,†} Yanan Yu,^{8,†} Dongmei Chen,^{9,†} Peng Zhou,^{1,†} Yin hao Wang,¹  Mingzhou Zhang,¹ Aimin Jiang,⁵ Zhimin Chen³ and Xuemin Li¹ 

¹Peking University Third Hospital, Beijing, China

²Beijing Fengtai Hospital, Beijing, China

³Hebei Eye Hospital, Xingtai, China

⁴General Hospital of Huabei Petroleum Administration Bureau, Renqiu, China

⁵The Hospital of Shunyi District, Beijing, China

⁶Datong Aier Eye Hospital, Datong, China

⁷People's Hospital of Yongqing, Langfang, China

⁸Baoding Zhuozhou Gem Flower Hospital, Zhuozhou, China

⁹Baoding Gem Flower Dongfang Hospital, Xushui, China

ABSTRACT.

Purpose: The present study aimed to investigate and compare the psychometric properties of the National Eye Institute Visual Functioning Questionnaire-25 (NEI VFQ-25) and the Visual Function Index-14 (VF-14) in a large sample of patients with cataracts.

Methods: A total of 1052 patients with bilateral age-related cataracts were recruited in the study. Patients with other comorbidities that severely impacted vision were excluded. Participants completed the two questionnaires in random order. Classical test theory and Rasch analyses were used to assess the psychometric properties of the questionnaires.

Results: Complete data were obtained from 899 patients. The mean overall index score on the NEI VFQ-25 was 76.1 ± 19.0 , while that on the VF-14 was 46.5 ± 15.0 . Cronbach's α -values for the NEI VFQ-25 and VF-14 were 0.89 and 0.95, respectively. Ceiling effects were observed on nine of the 12 subscales in the NEI VFQ-25. The correlation between total scores on the NEI VFQ-25 and VF-14 was moderate ($r = 0.600$; $p < 0.001$), and subscales of the NEI VFQ-25 were weakly or moderately correlated with the similar domains on the VF-14. Rasch analysis revealed ordered category thresholds and sufficient person separation for both instruments, while the two questionnaires had critical deficiencies in unidimensionality, targeting and differential item functioning.

Conclusion: Neither the NEI VFQ-25 nor VF-14 is optimal for the assessment of vision-related quality of life in typical Chinese patients with cataracts. The potential deficiencies of the questionnaires should be taken into consideration prior to application of the instruments or interpretation of the results.

Key words: cataract – NEI VFQ-25 – psychometric properties – Rasch analysis – VF-14 – vision-related quality of life

[†]These authors contributed equally to the work.

Acta Ophthalmol. 2021; 99: e480–e488

©2020 The Authors. Acta Ophthalmologica published by John Wiley & Sons Ltd on behalf of Acta Ophthalmologica Scandinavica Foundation

This is an open access article under the terms of the Creative Commons Attribution-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited and no modifications or adaptations are made.

doi: 10.1111/aos.14606

Introduction

Age-related cataract is one of the most common potential causes of blindness in older adults, which can significantly impact visual function and quality of life (Liu et al. 2017). Traditionally, visual function has been evaluated in these patients using assessments of visual acuity (e.g. Snellen chart) (Sparrow et al. 2018). However, as visual impairments caused by cataracts lead to a wide range of functional problems in the daily life, clinical measures of visual acuity do not sufficiently reflect the overall visual function in those with cataracts (Groessl et al. 2013). Self-reported measures of vision-related quality of life allow one to examine a patient's subjective well-being and visual disability while performing daily tasks (Fung et al. 2016). In recent years, numerous questionnaires regarding vision-related quality of life have been developed and applied in both clinical and research settings (Massof & Rubin 2001; Rentz et al. 2014).

The National Eye Institute Visual Functioning Questionnaire-25 (NEI VFQ-25) and the Visual Function

Index-14 (VF-14) are two commonly used questionnaires regarding visual function (Chiang et al. 2011). The NEI VFQ-25 is a multidimensional questionnaire designed to assess the impact of eye conditions/visual problems on general quality of life using 25 items across 12 subscales (Marrella et al. 2010). The VF-14 is a 14-item self-report questionnaire developed to measure functional impairment in patients with cataracts across 14 vision-dependent activities (Khadka et al. 2014). Both questionnaires have been translated into a number of different languages for widespread use in clinical and research settings (Valderas et al. 2005; To et al. 2014; Ni et al. 2015; Pan et al. 2015; Zhu et al. 2015).

To ensure the accuracy and effectiveness of measurement, an instrument should exhibit satisfactory psychometric properties. Classical test theory (CTT) is a conventional quantitative approach for investigating the reliability and validity of a scale (Cappelleri et al. 2014). CTT is based on the assumption that all items are equal indicators, focusing on the questionnaire as a whole (Petrillo et al. 2015). In contrast, Rasch analysis is a type of modern test theory which converts ordinal rating scale observations into continuum probability measures (Skiadaresi et al. 2016), allowing one to investigate both person ability and item difficulty on the same metric scale (Xu et al. 2018). Before applying an instrument to a specific population, it is crucial to systematically evaluate its psychometric properties within that population. Researchers should also remain aware of whether similar instruments measure the same construct—a premise necessary for making comparisons across studies despite the use of different instruments (Gothwal et al. 2012). Although several studies have separately investigated the NEI VFQ-25 and VF-14, none have directly compared these two instruments in the same patient population. The present large-scale, multicentre study aimed to investigate the psychometric properties of the NEI VFQ-25 and VF-14 in typical Chinese patients with cataracts and to determine whether these assessments measure the same construct.

Methods

Study cohort

This study was conducted across eight medical centres in China from January

to June of 2019. A total of 1052 prospective patients with bilateral age-related cataracts were enrolled in the study. Patients with severe comorbidities affecting vision (e.g. macular diseases, glaucoma, optic atrophy and proliferative diabetic retinopathy) were excluded. Lens opacities were graded by well-trained ophthalmologists in each medical centre via slit-lamp examination in accordance with the standard colour figures of the Lens Opacities Classification System III (LOCS III) (Chylack et al. 1993). Binocular uncorrected distance visual acuity (UDVA) values were also recorded for each participant. During administration, participants completed the two questionnaires in random order. For subjects who cannot figure out or understand the questionnaires well by themselves, the ophthalmologists would explain the questions to them to help understand.

This study adhered to the tenets outlined in the Declaration of Helsinki. Ethical approval was obtained from each local institutional review board, and written informed consent was obtained from each participant.

Questionnaires

National Eye Institute Visual Functioning Questionnaire-25

The NEI VFQ-25 is composed of 25 items designed to assess visual disability and health-related quality of life (Hirneiss et al. 2010). The questionnaire contains a single general health construct and 11 visual functioning subscales: general vision, ocular pain, colour vision, near activities, distance activities, social function, mental health, role difficulties, dependency, driving and peripheral vision (Nickels et al. 2017). In total, there are six question types in the NEI VFQ-25 (global rating of health, difficulty, frequency, severity, agreement and numeric) with different response categories (numeric with no labels, all of the time—none of the time, definitely true—definitely false, none—very severe, excellent—completely blind and no difficulty at all—stopped doing this because of your eyesight) (Mollazadegan et al. 2014). For difficulty-scale questions, ‘stopped doing this for other reasons or not interested in doing this’ responses are treated as missing data. Items within each subscale are

converted to a subscale score ranging from 0 to 100, and the overall composite score is calculated by averaging the 11 vision-targeted subscale scores. A higher score indicates better vision-specific quality of life (Kay & Ferreira 2014).

Visual Function Index-14

The VF-14 contains 14 questions rating the degree of difficulty in performing 14 vision-targeted activities of daily living (i.e. reading small print, reading newspaper, reading large print, recognizing people, seeing stairs, reading traffic signs, doing fine handwork, writing checks, playing games, taking part in sports, preparing meals, watching television, driving in daylight and driving at night) (Las Hayas et al. 2011b). Responses range from ‘no difficulty’ to ‘unable to do the activity’ across five response categories. ‘Not applicable’ responses are treated as missing data in the analysis (Gothwal et al. 2010). Scores are calculated as the summated responses divided by the number of valid responses. The value is then multiplied by 25 to yield a final score ranging between 0 and 100, with 0 representing the worst and 100 representing the best possible functioning (Las Hayas et al. 2011a).

Statistical analyses

SPSS version 23.0 (IBM, Chicago, IL, United States) and Winsteps version 3.72.3 (Winsteps, Chicago, IL, United States) were used for statistical analysis. A p-value of less than 0.05 was considered statistically significant.

Classical test theory

The internal consistency/reliability of each questionnaire was assessed using Cronbach’s α coefficients. The acceptable minimum value of Cronbach’s α is 0.70, and a coefficient of ≥ 0.80 is considered indicative of good reliability (Wang et al. 2008). For the NEI VFQ-25, the average item–total score correlation for each multi-item subscale was also calculated using Spearman’s correlation coefficients. Spearman correlation coefficients < 0.25 are considered to indicate a weak or absent relationship. Values ranging from 0.25 to 0.50 indicate a fair relationship, those ranging from 0.50 to 0.75 indicate a moderate-to-good relationship, and those > 0.75 indicate an excellent relationship (Kovac et al. 2015). Floor or ceiling effects

are considered to be present if more than 15% of responses achieve the minimum or maximum score, respectively (Terwee et al. 2007). If floor or ceiling effects are present, it is likely that extreme items are missing in the lower or upper end of the scale, indicating limited content validity. As a consequence, patients with the lowest or highest possible score cannot be distinguished from each other; thus, reliability is reduced (Terwee et al. 2007). In addition, we examined concurrent validity by assessing correlations between responses on the NEI VFQ-25 and VF-14 using Spearman's correlation analyses.

Rasch analysis

In addition to the traditional methods described above, we performed Rasch analysis to further evaluate the psychometric properties of the NEI VFQ-25 and VF-14. The ranking of response categories was reversed when necessary to guarantee that higher scores always represented higher levels of visual functioning. Rasch analysis was conducted using the Andrich Rating Scale Model (Andrich 1978). The following components were examined: category threshold order, item fitness, principal component analysis, person separation, targeting and differential item functioning.

Category threshold order. The ordering of the response category threshold is an essential parameter that reflects the usage of response categories by the respondents. Category probability curve graphs were generated to display the likelihood of each category being selected over the range of measurement (Khadka et al. 2012). In a well-functioning rating scale, thresholds should be arranged in a hierarchical order, with each curve showing a distinct peak. Disordered thresholds indicate that the response categories are underused or difficult to distinguish. As disordered thresholds can be a cause of item misfit, dysfunctional rating scales are required to be modified prior to further analysis.

Item fitness. Item fitness is used to examine the unidimensionality of a psychometric measure, indicating how closely each item response matches the expectations of the Rasch model. There are two types of fit statistics: information-weighted (infit) and outlier-sensitive (outfit). The infit statistic assesses adverse reactions to the items at the competency level of the participant, while the outfit statistic assesses adverse reactions to the items far from the participant's ability level. Both item

fit statistics are expressed as the mean square (MnSq), with an expected value of 1.0. Though the stringent criterion for infit and outfit MnSq statistics is between 0.7 and 1.3, a more lenient criterion between 0.5 and 1.5 is also considered acceptable (Pesudovs et al. 2007; Janssen et al. 2017).

Principal component analysis. Principal component analysis (PCA) of the residuals is also a tool for determining the unidimensionality of a questionnaire. For a unidimensional instrument, most of the variance should be explained by the principal component, with no evidence of a second construct being measured (Khadka et al. 2014). To be considered unidimensional, the variance explained by the principal component for the empirical calculation should be comparable to that explained by the model and should be >60%, while the variance unexplained by the contrast should be <2.0 eigenvalue units (Smith 2002).

Person separation. The person separation index is a measure of an instrument's precision and is used to estimate how broadly the participants can be discriminated into statistically distinct levels. The person separation reliability coefficient describes the reliability of the scale to discriminate between individuals of different abilities. A person separation index of ≥ 2.0 and a person separation reliability of ≥ 0.8 represent the minimum acceptable level of separation (Marella et al. 2010).

Targeting. Targeting refers to how well the difficulty of items in the scale matches the abilities of the individuals in the sample. Targeting can be evaluated by visually inspecting person-item maps and measuring the difference between person and item mean values. Targeting is indicated when the person and item means are close to each other. A discrepancy between means of >1.0 logit indicates notable mistargeting (Pesudovs et al. 2007).

Differential item functioning. Differential item functioning (DIF) assesses whether the items function similarly at the same level of ability regardless of the respondent's demographic characteristics. In the present study, the DIF of each item was assessed according to sex, age (≤ 70 years and >70 years), and systemic comorbidity (present and absent). DIF was considered absent when the difference in the item measure was <0.50 logits, minimal when the difference ranged from 0.5 to 1.0 logits,

and notable when the difference was >1.0 logit (Khadka et al. 2011).

Results

Demographic and clinical characteristics of the study participants

Among the 1052 patients recruited for the study, 153 patients were excluded due to incomplete clinical data or improper answers (e.g. choosing multiple options in a single question). Complete data were obtained from 899 patients (85.5%), all of whom were included in the final analysis. The study cohort comprised 364 men (40.5%) and 535 women (59.5%). The mean age of the participants was 70.3 ± 9.8 years (mean \pm SD), ranging from 40 to 90 years. Mean UDVA values were 0.48 ± 0.31 LogMAR in the better eye and 0.76 ± 0.37 LogMAR in the worse eye. Mean LOCS III grading scores were 2.77 ± 0.80 (NO), 2.45 ± 0.98 (C), and 2.22 ± 1.20 (P), respectively. The detailed demographic and clinical characteristics of the study cohort are presented in Table 1.

Psychometric properties of the VF-14 and NEI VFQ-25

The mean overall index score on the VF-14 was 46.5 ± 15.0 , while the mean

Table 1. Demographic and clinical characteristics of the study cohort.

Variables	Sample, n = 899
Age (years)	
Mean \pm SD	70.3 \pm 9.8
Range	40–90
Gender, number (%)	
Male	364 (40.5)
Female	535 (59.5)
UDVA (LogMAR), mean \pm SD	
Better eye	0.48 \pm 0.31
Worse eye	0.76 \pm 0.37
LOCS III score, mean \pm SD	
NO	2.77 \pm 0.80
C	2.45 \pm 0.98
P	2.22 \pm 1.20
Systemic comorbidity, number (%)	
Hypertension	385 (42.8)
Diabetes	205 (22.8)
Heart disease	110 (12.2)
Cerebrovascular disease	35 (3.9)

C = cortical, LOCS III = Lens Opacities Classification System III, NO = nuclear opalescence, P = posterior subcapsular, SD = standard deviation, UDVA = uncorrected distance visual acuity.

Table 2. Reliability and validity analysis of the NEI VFQ-25 and VF-14 using classical test theory.

Index (item number of subscale)	Score (mean ± SD)	Cronbach's α	Average item–total score correlation	Floor responses (%)	Ceiling responses (%)
VF-14	46.5 ± 15.0	0.95	0.709	0.9	0.1
NEI VFQ-25	76.1 ± 19.0	0.89	0.734	0.0	0.1
General health (1)	39.4 ± 19.7	NA	NA	6.6	1.0
General vision (1)	45.5 ± 16.8	NA	NA	0.2	0.1
Ocular pain (2)	85.5 ± 20.8	0.67	0.860	0.3	56.3
Near activities (3)	67.5 ± 24.7	0.80	0.738	1.4	14.2
Distance activities (3)	77.6 ± 23.7	0.90	0.762	0.8	36.9
Social functioning (2)	86.8 ± 22.3	0.89	0.823	1.1	63.5
Mental health (4)	78.7 ± 24.9	0.88	0.772	0.9	35.2
Role difficulties (2)	71.8 ± 29.5	0.93	0.932	1.7	44.6
Dependency (3)	81.0 ± 27.2	0.93	0.942	1.1	59.0
Driving* (3)	73.4 ± 35.1	0.77	0.744	12.2	49.1
Colour vision (1)	88.4 ± 22.8	NA	NA	1.7	73.0
Peripheral vision (1)	81.1 ± 23.7	NA	NA	0.6	51.5

NA = not applicable (statistics need two or more items), NEI VFQ-25 = National Eye Institute Visual Functioning Questionnaire, SD = standard deviation, VF-14 = Visual Function Index.

* 287 samples in the 'Driving' subscale due to missing responses (not driving for other reasons).

total NEI VFQ-25 score was 76.1 ± 19.0. Mean subscale scores for the NEI VFQ-25 are presented in Table 2. A high missing response rate was detected in the 'driving' subscale of the NEI VFQ-25 (missing percentages of 68.0% and 76.4% for items 15 and 16, respectively). Similarly, we also observed a high missing response rate for VF-14 items related to driving (missing percentages of 67.3% and 75.6% for items 13 and 14, respectively).

Classical test theory

Cronbach's α values for the VF-14 and NEI VFQ-25 were 0.95 and 0.89, respectively, indicating good internal consistency for each questionnaire. Cronbach's α values for the multi-item subscales of the NEI VFQ-25 ranged from 0.67 to 0.93 (Table 2). Overall, reliability was satisfactory for the majority of subscales, although one subscale exhibited a Cronbach's α below 0.70 ('ocular pain', 0.67). Our analysis of average item–total score correlations showed that the items which reflected the same construct yielded similar results (Table 2).

Although we observed no floor or ceiling effect for the VF-14, ceiling effects were identified in nine of the 12 subscales of the NEI VFQ-25. The percentage of patients with the maximum score ranged from 35.2% ('mental health' subscale) to 73.0% ('colour vision' subscale; Table 2). No subscale of the NEI VFQ-25 was detected to have a floor effect.

In our analysis of concurrent validity, Spearman correlation coefficients revealed a moderate correlation between total scores on the VF-14 and NEI VFQ-25 ($r = 0.600$; $p < 0.001$). Subscales on the NEI VFQ-25 were weakly or moderately correlated with similar domains on the VF-14 (Table 3).

Rasch analysis

Appropriateness of rating scale. The category threshold order for the NEI VFQ-25 and VF-14 is illustrated in Fig. 1. All category probability curves intersected and increased in an orderly fashion, suggesting that all categories were utilized by the respondents and that the rating scales are appropriate for each assessment.

Item fitness. The fit statistics for each item on the NEI VFQ-25 or VF-14 are presented in Table 4. For the NEI VFQ-25, infit values ranged from 0.47 to 2.33, while outfit values ranged from 0.42 to 2.15. Thirteen misfitting items were identified: general health, general vision, pain around the eyes, reading normal newsprint, seeing well up close, finding objects on a crowded shelf, reading street signs, seeing objects off to the side, matching clothes, visiting others, driving during daytime, driving at night, and driving in difficult conditions. For the VF-14, infit values ranged from 0.58 to 1.91, while outfit values ranged from 0.56 to 1.96, with three misfitting items (reading small print, driving in daylight, and driving at night). The results suggest that both

instruments contained items that introduced noise into the data and did not measure the underlying construct.

Principal component analysis. For the NEI VFQ-25, PCA of item residuals revealed that the variance explained by measures for the empirical calculation (59.2%) was comparable to that of the model (61.8%). The unexplained variance in the first contrast was 4.7 eigenvalue units, while that in the second contrast was 2.8 eigenvalue units. No further contrasts exceeded 2.0 eigenvalue units (Table 5). Similarly, the PCA of item residuals for the VF-14 also revealed that the variance explained by the principal component was comparable for the empirical calculation (57.8%) and the model (57.7%). The unexplained variance in the first contrast accounted for 4.2 eigenvalue units, with no further contrasts exceeding 2.0 eigenvalue units. These findings suggest that the questionnaires were not unidimensional and indicate the presence of a second dimension on each scale.

Person separation. The separation index for person measures of the NEI VFQ-25 was 3.38, with a reliability of 0.92. The person separation index for the VF-14 was 3.08, with a reliability of 0.90 (Table 5). These results suggest excellent discrimination between individuals of different abilities by both questionnaires.

Targeting. The person–item maps are displayed in Fig. 2. In each person–item map, the left half of the graph represents the distribution of persons,

Table 3. Correlations between the VF-14 index and the NEI VFQ-25 subscales.

	General health	General vision	Ocular pain	Near activities	Distance activities	Social functioning	Mental health	Role difficulties	Dependency	Driving	Colour vision	Peripheral vision	NEI VFQ-25 total
Reading small print	0.236**	0.234**	0.259**	0.334**	0.272**	0.274**	0.093**	0.318**	0.176**	0.232**	0.242**	0.249**	0.326**
Reading newspaper	0.217**	0.171**	0.318**	0.455**	0.379**	0.378**	0.235**	0.413**	0.303**	0.220**	0.342**	0.360**	0.434**
Reading large print	0.168**	0.200**	0.277**	0.331**	0.464**	0.422**	0.353**	0.394**	0.349**	0.284**	0.373**	0.446**	0.448**
Recognizing people	0.072**	0.101**	0.138**	0.256**	0.255**	0.227**	0.239**	0.228**	0.221**	0.221**	0.230**	0.226**	0.263**
Seeing stairs	0.091**	0.199**	0.229**	0.283**	0.390**	0.339**	0.267**	0.322**	0.241**	0.358**	0.311**	0.346**	0.361**
Reading traffic signs	0.141**	0.238**	0.306**	0.265**	0.491**	0.435**	0.375**	0.425**	0.345**	0.317**	0.415**	0.472**	0.475**
Doing fine handwork	0.206**	0.203**	0.379**	0.420**	0.400**	0.417**	0.243**	0.421**	0.305**	0.297**	0.391**	0.390**	0.463**
Writing checks	0.206**	0.108**	0.506**	0.433**	0.571**	0.566**	0.511**	0.518**	0.523**	0.298**	0.547**	0.566**	0.600**
Playing games	0.191**	0.188**	0.389**	0.349**	0.509**	0.502**	0.385**	0.464**	0.405**	0.315**	0.473**	0.508**	0.511**
Taking part in sports	0.109**	0.133**	0.295**	0.280**	0.398**	0.402**	0.320**	0.345**	0.327**	0.269**	0.374**	0.379**	0.390**
Preparing meals	0.203**	0.242**	0.296**	0.264**	0.404**	0.397**	0.247**	0.377**	0.268**	0.211**	0.341**	0.365**	0.390**
Watching TV	0.184**	0.231**	0.277**	0.283**	0.435**	0.342**	0.264**	0.399**	0.261**	0.335**	0.290**	0.400**	0.401**
Driving in daylight	0.205**	0.196**	0.383**	0.514**	0.539**	0.471**	0.436**	0.474**	0.444**	0.390**	0.461**	0.469**	0.575**
Driving at night	0.197**	0.148**	0.373**	0.480**	0.505**	0.419**	0.471**	0.486**	0.453**	0.379**	0.442**	0.458**	0.553**
VF-14 Total	0.208**	0.207**	0.464**	0.471**	0.594**	0.580**	0.428**	0.559**	0.452**	0.365**	0.531**	0.569**	0.600**

NEI VFQ-25 = National Eye Institute Visual Functioning Questionnaire, VF-14 = Visual Function Index.

** Significant at the 0.01 level (two-tailed).

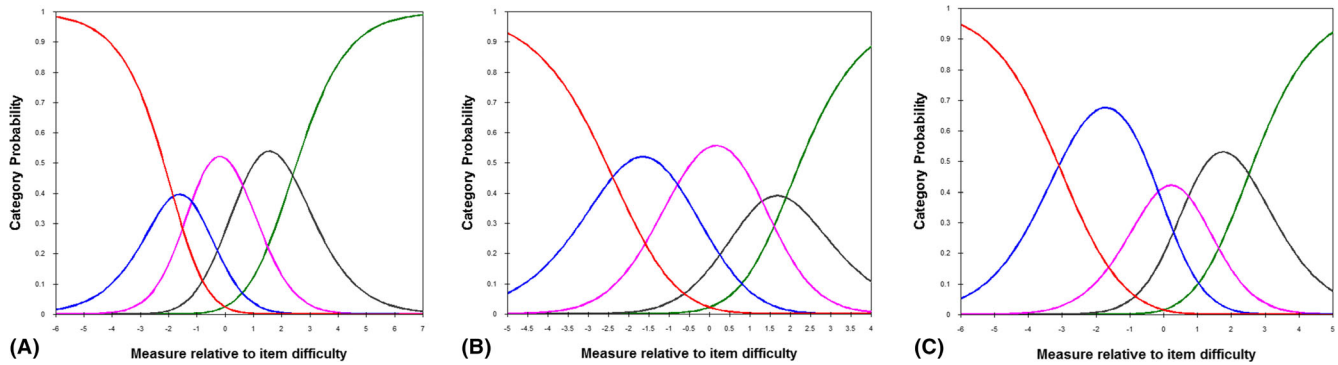


Fig. 1. Category probability curves. The x-axis of the category probability curve reflects the personal latent characteristic and the logit difference of item difficulty, while the y-axis reflects the probability of the category being chosen. (A) Category probability curves for the questions with difficulty-related response options in the National Eye Institute Visual Functioning Questionnaire-25 (NEI VFQ-25). The category probability curve was intersecting and ordered. The curve at the extreme left represents ‘stopped doing this because of eyesight’, while the curve at the extreme right represents ‘no difficulty at all’. The response option ‘stopped doing this for other reasons or not interested in doing this’ was treated as missing data and not included in the analysis. (B) Category probability curves for questions with frequency-related response options in the NEI VFQ-25, illustrating ordered thresholds. The five curves from left to right represent response categories ranging from ‘all of the time’ to ‘none of the time’. (C) Category probability curves for questions with agreement-related response options in the NEI VFQ-25, illustrating ordered thresholds. The five curves from left to right represent response categories ranging from ‘definitely true’ to ‘definitely false’. (D) Category probability curves for the Visual Function Index. Each response option was appropriately used, and category calibration increased in an ordered fashion. The curve at the extreme left represents ‘unable to do’, while the curve at the extreme right represents ‘no difficulty’. The response option ‘not applicable’ was treated as missing data.

Table 4. Fit Statistics of the NEI VFQ-25 and VF-14 using Rasch analysis.

Items of the NEI VFQ-25	Measure	Infit MnSq	Outfit MnSq	Items of the VF-14	Measure	Infit MnSq	Outfit MnSq
General health	2.20	1.23	1.62	Reading small print	1.13	1.89	1.96
General vision	1.25	1.61	2.15	Reading newspaper	0.57	1.21	1.25
Worry about eyesight	0.38	1.31	1.30	Reading large print	-0.58	0.68	0.69
Pain around the eyes	-0.86	1.33	1.54	Recognizing people	-1.21	1.17	1.15
Reading normal newsprint	1.05	1.65	1.50	Seeing stairs	-0.75	0.74	0.73
Seeing well up close	0.88	1.54	1.38	Reading traffic signs	-0.27	0.71	0.71
Finding objects on a crowded shelf	-0.37	0.56	0.49	Doing fine handwork	0.84	1.40	1.43
Reading street signs	-0.30	0.50	0.45	Writing checks	-0.62	0.75	0.72
Going downstairs at night	0.07	0.76	0.67	Playing games	-0.78	0.58	0.56
Seeing objects off to the side	-0.35	0.47	0.42	Taking part in sports	-0.94	0.74	0.72
Seeing how people react	-0.91	0.71	0.52	Preparing meals	-1.15	0.65	0.62
Matching clothes	-1.03	0.85	0.48	Watching TV	-0.39	0.78	0.78
Visiting others	-0.88	0.69	0.45	Driving in daylight	1.68	1.91	1.80
Going out to movies/plays	-0.06	0.80	0.68	Driving at night	2.47	1.71	1.59
Driving during daytime	-0.21	2.33	1.70				
Driving at night	0.76	2.04	1.57				
Driving in difficult conditions	0.70	2.03	1.53				
Accomplish less	0.29	0.92	0.82				
Limited endurance	0.34	0.84	0.73				
Amount of time in pain	-0.66	1.15	0.76				
Stay home most of the time	-0.33	0.95	0.70				
Frustrated	-0.38	1.05	0.84				
No control	-0.34	1.01	0.73				
Rely too much on others’ words	-0.39	0.94	0.64				
Need much help from others	-0.35	0.94	0.65				
Embarrassment	-0.49	0.83	0.54				

MnSq = mean square, NEI VFQ-25 = National Eye Institute Visual Functioning Questionnaire-25, VF-14 = Visual Function Index.

while the right half represents the distribution of the item measures. In the person-item map for the NEI VFQ-25, the person and item means were far apart. Most of the items covered people with low and moderate visual ability, while a considerable portion of the persons with high visual

ability was uncovered, indicating notable mistargeting of the NEI VFQ-25. The person-item map for the VF-14 exhibited better targeting of item difficulty to patient ability than the NEI VFQ-25, given that the ‘M’ values for person and item were closer together. In contrast to those on the NEI VFQ-

25, items on the VF-14 were more inclined to cover people with moderate and high visual ability.

Differential item functioning. For the NEI VFQ-25, one item exhibited minimal DIF based on sex (driving in difficult conditions, logit value: 0.51), while three items exhibited DIF based

Table 5. Comparison of overall performance in Rasch analysis between the NEI VFQ-25 and VF-14.

Components	NEI VFQ-25	VF-14
Unidimensionality		
Misfitting items (<i>n</i>)	13	3
Variance explained by measures (empirical)	59.2%	57.8%
Variance explained by measures (modelled)	61.8%	57.7%
Unexplained variance in 1st contrast	4.7	4.2
Unexplained variance in 2nd contrast	2.8	1.9
Unexplained variance in 3rd contrast	1.9	1.5
Unexplained variance in 4th contrast	1.8	1.2
Unexplained variance in 5th contrast	1.6	1.0
Person separation		
Person separation index	3.38	3.08
Person separation reliability	0.92	0.90
DIF		
By age (items, <i>n</i>)	0	0
By sex (items, <i>n</i>)	1	2
By systemic comorbidity (items, <i>n</i>)	3	2

DIF = differential item functioning, NEI VFQ-25 = National Eye Institute Visual Functioning Questionnaire-25, VF-14 = Visual Function Index.

on systemic comorbidity (general health, 0.57; seeing well up close, 0.51; matching clothes, 0.57; Table 5). For the VF-14, two items exhibited DIF based on sex (driving in daylight, 0.94; driving at night, 0.64), while an additional two exhibited DIF based on systemic comorbidity (recognizing people, 0.59; driving at night, 0.52). No item on either questionnaire exhibited notable DIF (>1.0 logit).

Discussion

In the present study, we investigated and compared the psychometric performance of the NEI VFQ-25 and VF-14 in a large sample of patients with cataracts across eight medical centres in China. Our results indicated that both questionnaires exhibited deficiencies in several critical measurement properties.

We first examined the reliability and validity of the instruments using CTT. Cronbach's α , a key measure of reliability, was high for the overall index values of either questionnaire (0.89 for the NEI VFQ-25 and 0.95 for the VF-14). However, one subscale of the NEI VFQ-25 exhibited a Cronbach's α value below 0.70 ('ocular pain', 0.67). Therefore, although the overall reliability of the questionnaire was quite satisfactory, high internal consistency was not ensured in every subscale. In regard to validity, most subscales of the NEI VFQ-25 exhibited a ceiling effect. Such effects have also been reported in

previous studies attempting to validate translated versions of the assessment (Labiris et al. 2008; Nickels et al. 2017). High ceiling percentages suggest reduced content validity and an inability to distinguish participants with high visual function. We also examined the concurrent validity of the two questionnaires by comparing the subscales of the NEI VFQ-25 with items of the VF-14. We observed weak-to-moderate correlations between the two instruments, indicating that the NEI VFQ-25 and VF-14 may not have reflected the same construct in our population. Therefore, caution should be exercised when making comparisons of vision-specific quality of life across studies using different instruments.

We also performed Rasch analysis to assess the psychometric properties of the two questionnaires. Rasch analysis is a modern psychometric validation technique that can provide a deeper understanding of various critical measurement properties of an instrument (Leung et al. 2008). Our Rasch analysis revealed ordered category thresholds in the NEI VFQ-25 and VF-14, indicating that our participants understood the content of the categories well and distinguished between the characteristics of the categories correctly. We also observed sufficient person separation of the instruments, suggesting that both questionnaires can distinguish among different levels of ability. However, several crucial deficiencies were also detected, including multidimensionality, mistargeting and DIF. Item fitness and

PCA of the residuals are two key indicators used to assess the unidimensionality of an instrument. Misfitting was observed for thirteen of the 25 NEI VFQ-25 items and three of the 14 VF-14 items. While previous studies have also reported multidimensionality for the two instruments, the misfit items were not quite the same (Valderas et al. 2004; Pesudovs et al. 2010). PCA of the residuals also suggested the presence of a second dimension in both scales. Multidimensionality represents a significant problem, as it invalidates the reporting of a total score derived from all items. Targeting measures how well the difficulty of items in the scale matches the abilities of the persons in the sample. In accordance with the findings of previous studies (Gothwal et al. 2012; Kovac et al. 2015), the person-item maps of both instruments exhibited problems such as measurement redundancies and measurement voids. For example, the majority of items in the NEI VFQ-25 were located at similar difficulty regions, especially at the bottom of the continuum. Meanwhile, there was a shortage of items at the top. Such findings indicate that the instrument had a low ability to differentiate among participants with higher levels of visual function. The notable mistargeting of the NEI VFQ-25 may result from its severe ceiling effect. Although maps were somewhat better for the VF-14, the same general problems were observed. In contrast to findings observed for the NEI VFQ-25, measurement voids for the VF-14 were observed at the bottom of the continuum, suggesting a weakness of the instrument to distinguish participants with worse visual performance. Furthermore, the NEI VFQ-25 and VF-14 were both limited by notable DIF for sex and systemic comorbidity across several items, suggesting that subgroups of participants with comparable levels of ability responded differently to the items. Thus, bias may have been introduced by characteristics other than item difficulty in our population.

Previous studies have attempted to improve the psychometric properties of the scales by removing the misfitting items using Rasch analysis (Gothwal et al. 2010; Kovac et al. 2015; Abe et al. 2019). However, reengineering the questionnaires by simply reducing the number of misfitting items did not achieve satisfactory psychometric properties in our study. Developing and adding new items seemed to be helpful, as there were notable

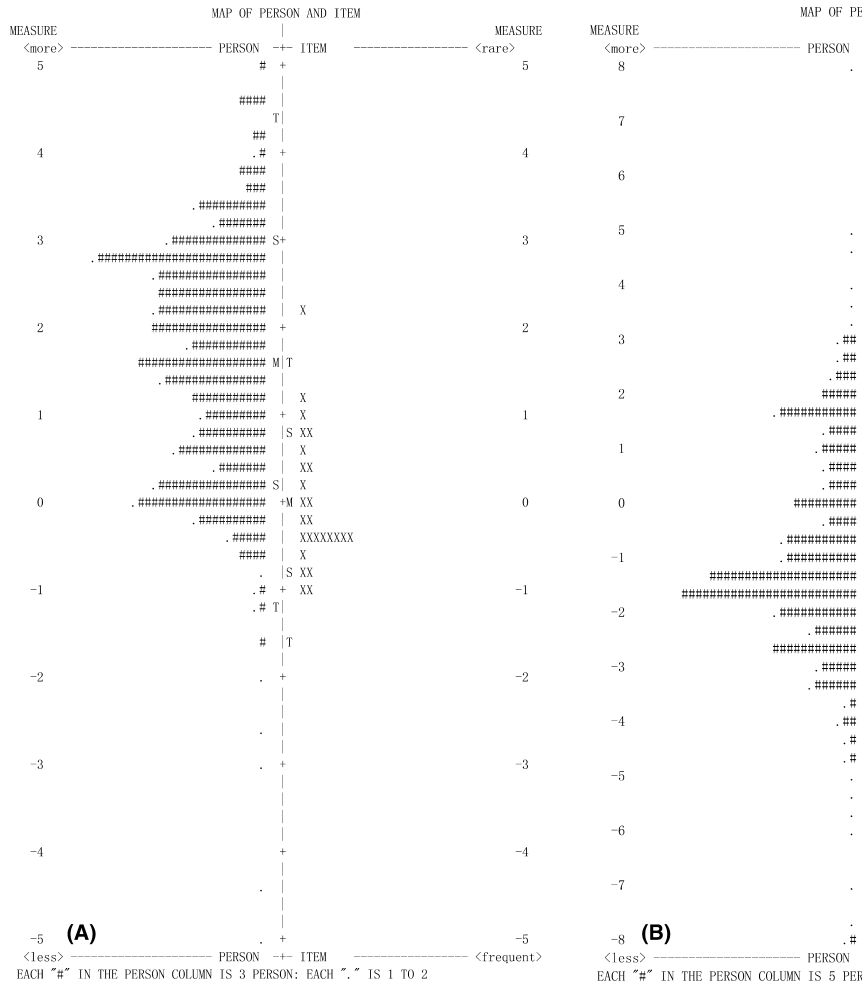


Fig. 2. Person–item maps for the National Eye Institute Visual Functioning Questionnaire-25 (NEI VFQ-25) (A) and Visual Function Index (B). In each person–item map, participants are displayed to the left of the dashed line, with more able participants located at the top of the map. Items are located to the right of the dashed line, with more difficult items located at the top of the map. The vertical line represents the measure of the variable in linear logit units. ‘M’ markers represent the location of the mean measure. ‘S’ markers are placed one sample standard deviation away from the mean. ‘T’ markers are placed two sample standard deviations away.

differences between the range of participant competence and the range of item difficulty (as indicated by the notable measurement voids in the person–item maps). Our findings suggest that it is necessary to add items with high difficulty to the NEI VFQ-25 and items with low difficulty to the VF-14.

Taken together, our findings highlight the potential deficiencies of the NEI VFQ-25 and VF-14. Thus, caution should be exercised when applying the assessments to specific populations. Despite modifications designed to address the suboptimal properties of these instruments, the NEI VFQ-25 and VF-14 are still widely used in their original formats in numerous ophthalmology clinics and research studies (Mollazadegan et al. 2014). In some ways, this is reasonable, as even slightly

different versions of the same questionnaire may add confusion and increase the difficulty of making comparisons among studies. A deeper understanding of the psychometric properties of the instruments, especially of their deficiencies and weaknesses, can help researchers and clinicians to properly interpret the results of their work. Notably, our results revealed the psychometric deficiencies of the instruments within a typical cataract population. Further studies are required to validate the questionnaires in populations with other ocular impairments.

The present study also had two limitations of note. First, due to the cross-sectional nature of our study, we were unable to examine the test–retest reliability of the assessments. Second, given that we aimed to investigate the

psychometric performance of the two instruments in a typical cataract population, all participants in our study had been diagnosed with cataracts. As the NEI VFQ-25 is designed to measure functional impairment associated with multiple visual problems, our results regarding the NEI VFQ-25 may not be applicable to the full clinical spectrum of ocular disease.

Conclusion

Our findings demonstrated deficiencies in the NEI VFQ-25 and VF-14 with regard to several critical psychometric properties, suggesting that neither is optimal for assessing vision-related quality of life in typical Chinese patients with cataracts. Further studies are required to validate these two assessments in more diverse populations.

References

Abe RY, Medeiros FA, Davi MA et al. (2019): Psychometric properties of the Portuguese version of the National Eye Institute Visual Function Questionnaire-25. *PLoS One* **14**: e0226086.

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

Cappelleri JC, Jason Lundy J & Hays RD (2014): Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther* **36**: 648–662.

Chiang PP, Fenwick E, Marella M, Finger R & Lamoureux E (2011): Validation and reliability of the VF-14 questionnaire in a German population. *Invest Ophthalmol Vis Sci* **52**: 8919–8926.

Chylack LT Jr, Wolfe JK, Singer DM et al. (1993): The Lens Opacities Classification System III. The Longitudinal Study of Cataract Study Group. *Arch Ophthalmol (Chicago, Ill : 1960)* **111**: 831–836.

Fung SS, Luis J, Hussain B, Bunce C, Hingorani M & Hancox J (2016): Patient-reported outcome measuring tools in cataract surgery: clinical comparison at a tertiary hospital. *J Cataract Refract Surg* **42**: 1759–1767.

Gothwal VK, Reddy SP, Sumalini R, Bharani S & Bagga DK (2012): National Eye Institute Visual Function Questionnaire or Indian Vision Function Questionnaire for visually impaired: a conundrum. *Invest Ophthalmol Vis Sci* **53**: 4730–4738.

Gothwal VK, Wright TA, Lamoureux EL & Pesudovs K (2010): Measuring outcomes of cataract surgery using the Visual Function Index-14. *J Cataract Refract Surg* **36**: 1181–1188.

Groessl EJ, Liu L, Sklar M, Tally SR, Kaplan RM & Ganiats TG (2013): Measuring the

- impact of cataract surgery on generic and vision-specific quality of life. *Qual Life Res* **22**: 1405–1414.
- Hirneiss C, Schmid-Tannwald C, Kernt M, Kampik A & Neubauer AS (2010): The NEI VFQ-25 vision-related quality of life and prevalence of eye disease in a working population. *Graefes Arch Clin Exp Ophthalmol* **248**: 85–92.
- Janssen KC, Phillipson S, O'Connor J & Johns MW (2017): Validation of the epworth sleepiness scale for children and adolescents using Rasch analysis. *Sleep Med* **33**: 30–35.
- Kay S & Ferreira A (2014): Mapping the 25-item National Eye Institute Visual Functioning Questionnaire (NEI VFQ-25) to EQ-5D utility scores. *Ophthalmic Epidemiol* **21**: 66–78.
- Khadka J, Gothwal VK, McAlinden C, Lamoureux EL & Pesudovs K (2012): The importance of rating scales in measuring patient-reported outcomes. *Health Qual Life Outcomes* **10**: 80.
- Khadka J, Huang J, Mollazadegan K, Gao R, Chen H, Zhang S, Wang Q & Pesudovs K (2014): Translation, cultural adaptation, and Rasch analysis of the visual function (VF-14) questionnaire. *Invest Ophthalmol Vis Sci* **55**: 4413–4420.
- Khadka J, Pesudovs K, McAlinden C, Vogel M, Kernt M & Hirneiss C (2011): Reengineering the glaucoma quality of life-15 questionnaire with Rasch analysis. *Invest Ophthalmol Vis Sci* **52**: 6971–6977.
- Kovac B, Vukosavljevic M, Djokic Kovac J, Resan M, Trajkovic G, Jankovic J, Smiljanic M & Grgurevic A (2015): Validation and cross-cultural adaptation of the National Eye Institute Visual Function Questionnaire (NEI VFQ-25) in Serbian patients. *Health Qual Life Outcomes* **13**: 142.
- Labiris G, Katsanos A, Fanariotis M, Tsirouki T, Pefkianaki M, Chatzoulis D & Tsironi E (2008): Psychometric properties of the Greek version of the NEI-VFQ 25. *BMC ophthalmology* **8**: 4.
- Las Hayas C, Bilbao A, Quintana JM, Garcia S & Lafuente I (2011a): A comparison of standard scoring versus Rasch scoring of the visual function index-14 in patients with cataracts. *Invest Ophthalmol Vis Sci* **52**: 4800–4807.
- Las Hayas C, Quintana JM, Bilbao A, Garcia S & Lafuente I (2011b): Visual acuity level, ocular morbidity, and the better seeing eye affect sensitivity and responsiveness of the visual function index. *Ophthalmology* **118**: 1303–1309.
- Leung YY, Tam LS, Kun EW, Ho KW & Li EK (2008): Comparison of 4 functional indexes in psoriatic arthritis with axial or peripheral disease subgroups using Rasch analyses. *J Rheumatol* **35**: 1613–1621.
- Liu YC, Wilkins M, Kim T, Malyugin B & Mehta JS (2017): Cataracts. *Lancet (London, England)* **390**: 600–612.
- Marella M, Pesudovs K, Keeffe JE, O'Connor PM, Rees G & Lamoureux EL (2010): The psychometric validity of the NEI VFQ-25 for use in a low-vision population. *Invest Ophthalmol Vis Sci* **51**: 2878–2884.
- Massof RW & Rubin GS (2001): Visual function assessment questionnaires. *Surv Ophthalmol* **45**: 531–548.
- Mollazadegan K, Huang J, Khadka J, Wang Q, Yang F, Gao R & Pesudovs K (2014): Cross-cultural validation of the National Eye Institute Visual Function Questionnaire. *J Cataract Refract Surg* **40**: 774–784.
- Ni W, Li X, Hou Z, Zhang H, Qiu W & Wang W (2015): Impact of cataract surgery on vision-related life performances: the usefulness of Real-Life Vision Test for cataract surgery outcomes evaluation. *Eye (London, England)* **29**: 1545–1554.
- Nickels S, Schuster AK, Singer S et al. (2017): The National Eye Institute 25-Item Visual Function Questionnaire (NEI VFQ-25) - reference data from the German population-based Gutenberg Health Study (GHS). *Health Qual Life Outcomes* **15**: 156.
- Pan AP, Wang QM, Huang F, Huang JH, Bao FJ & Yu AY (2015): Correlation among lens opacities classification system III grading, visual function index-14, pentacam nucleus staging, and objective scatter index for cataract assessment. *Am J Ophthalmol* **159**: 241–247.e242.
- Pesudovs K, Burr JM, Harley C & Elliott DB (2007): The development, assessment, and selection of questionnaires. *Optom Vis Sci* **84**: 663–674.
- Pesudovs K, Gothwal VK, Wright T & Lamoureux EL (2010): Remediating serious flaws in the National Eye Institute Visual Function Questionnaire. *J Cataract Refract Surg* **36**: 718–732.
- Petrillo J, Cano SJ, McLeod LD & Coon CD (2015): Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value Health* **18**: 25–34.
- Rentz AM, Kowalski JW, Walt JG et al. (2014): Development of a preference-based index from the National Eye Institute Visual Function Questionnaire-25. *JAMA Ophthalmol* **132**: 310–318.
- Skiafaresi E, Ravalico G, Polizzi S, Lundstrom M, Gonzalez-Andrades M & McAlinden C (2016): The Italian Catquest-9SF cataract questionnaire: translation, validation and application. *Eye Vis (London, England)* **3**: 12.
- Smith EV Jr (2002): Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* **3**: 205–231.
- Sparrow JM, Grzeda MT, Frost NA et al. (2018): Cataract surgery patient-reported outcome measures: a head-to-head comparison of the psychometric performance and patient acceptability of the Cat-PROM5 and Catquest-9SF self-report questionnaires. *Eye (London, England)* **32**: 788–795.
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM & de Vet HC (2007): Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* **60**: 34–42.
- To KG, Meuleners LB, Fraser ML et al. (2014): The impact of cataract surgery on vision-related quality of life for bilateral cataract patients in Ho Chi Minh City, Vietnam: a prospective study. *Health Qual Life Outcomes* **12**: 16.
- Valderas JM, Alonso J, Prieto L, Espallargues M & Castells X (2004): Content-based interpretation aids for health-related quality of life measures in clinical practice. An example for the visual function index (VF-14). *Qual Life Res* **13**: 35–44.
- Valderas JM, Rue M, Guyatt G & Alonso J (2005): The impact of the VF-14 index, a perceived visual function measure, in the routine management of cataract patients. *Qual Life Res* **14**: 1743–1753.
- Wang CW, Chan CL & Jin HY (2008): Psychometric properties of the Chinese version of the 25-item National Eye Institute Visual Function Questionnaire. *Optom Vis Sci* **85**: 1091–1099.
- Xu Z, Wu S, Li W, Dou Y & Wu Q (2018): The Chinese Catquest-9SF: validation and application in community screenings. *BMC Ophthalmol* **18**: 77.
- Zhu M, Yu J, Zhang J, Yan Q & Liu Y (2015): Evaluating vision-related quality of life in pre-operative age-related cataract patients and analyzing its influencing factors in China: a cross-sectional study. *BMC Ophthalmol* **15**: 160.

Received on March 15th, 2020.
Accepted on August 6th, 2020.

Correspondence:

Aimin Jiang
The Hospital of Shunyi District
No. 3 Guangming South Street
Shunyi District, Beijing
China
Tel: +86 15811186168
Email: jiang186168@163.com
Zhimin Chen
Hebei Eye Hospital
399 Quanbei East Street
Xingtai City, Hebei Province
China
Tel: +86 18932978678
Fax: +Xxxxx
Email: ykyyczm@126.com
Xuemin Li
Department of Ophthalmology
Peking University Third Hospital
49 North Garden Road
Haidian District, Beijing
China
Tel: +86 13911254862
Fax: +8601082089951
Email: 13911254862@163.com

†These authors contributed equally to the work.

This study was supported by the National Science and Technology Major Project (Grant Number: 2018ZX10101004).