

Translation, Cultural Adaptation, and Rasch Analysis of the Visual Function (VF-14) Questionnaire

Jyoti Khadka,^{1,2} Jinhai Huang,¹ Kaziwe Mollazadegan,³ Rongrong Gao,¹ Haisi Chen,¹ Sifang Zhang,¹ Qinmei Wang,¹ and Konrad Pesudovs²

¹School of Optometry and Ophthalmology and Eye Hospital, Wenzhou Medical University, Wenzhou, Zhejiang, China

²National Health and Medical Research Council Centre for Clinical Eye Research, Discipline of Optometry and Vision Science, Flinders University of South Australia, Bedford Park, South Australia, Australia

³St. Erik Eye Hospital, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

Correspondence: Jyoti Khadka, NH&MRC Centre for Clinical Eye Research, Discipline of Optometry and Vision Science, Flinders University of South Australia, Sturt South Wing (S149), Bedford Park, South Australia, 5042, Australia; jyoti.khadka@flinders.edu.au. Qinmei Wang, Eye Hospital of Wenzhou Medical University, 270 West Xueyuan Road, Wenzhou, Zhejiang, China, 325027; wqm6@mail.eye.ac.cn.

JK and JH are joint first authors. JK and JH contributed equally to the work presented here and should therefore be regarded as equivalent authors.

Submitted: January 22, 2014

Accepted: June 2, 2014

Citation: Khadka J, Huang J, Mollazadegan K, et al. Translation, cultural adaptation, and Rasch analysis of the Visual Function (VF-14) questionnaire. *Invest Ophthalmol Vis Sci*. 2014;55:4413–4420. DOI:10.1167/iovs.14-14017

PURPOSE. To translate, culturally adapt, and validate the original and previously validated shorter versions of the Visual Function Index (VF-14) questionnaire in a Chinese population.

METHODS. The VF-14 was completed by patients with cataract. The analysis was carried out in three phases: phase I, testing whether the VF-14 and its valid shorter versions, VF-8R and VF-11R, form valid scales in Chinese settings using Rasch analysis; phase II, developing completely new Chinese versions of the VF-14; phase III, testing whether the previously validated shorter versions of the VF-14 could be applied in a Chinese population. This was tested by assessing the agreement between the new Chinese (developed in phase II) and the previously validated shorter versions of the VF-14 using Bland-Altman plots.

RESULTS. A total of 456 patients (median age, 70 years; range, 40–92 years; females, 58%) completed the Chinese translated version of the VF-14. The VF-14 and the VF-11R demonstrated good Rasch based psychometric properties when a grossly misfitting item was removed. The VF-8R formed a valid scale without any modification. The scores of the VF-11R and the Chinese shorter version (VF-11RChin) showed very good agreement, with a mean difference of -0.18 logits and 95% limits of agreement between 0.11 and -0.47 .

CONCLUSIONS. The Chinese translated VF-14, VF-11R, and VF-8R were valid and could be applied to assess cataract outcomes in Chinese settings. The existing shorter version had good agreement with the new Chinese version, which signifies that there was no need to develop a different version of the VF-14 in China.

Keywords: patient-reported outcomes, VF-14, Rasch analysis, psychometric properties, cross-cultural validation

The Visual Function Index (VF-14) is a widely used patient-reported outcome (PRO) instrument in ophthalmology. The VF-14 was developed by Steinberg et al.¹ in 1994. The instrument was originally developed to assess vision-specific activity limitations in patients with cataract and to assess cataract surgical outcomes.¹ Since then, the instrument has been used to assess the impact and treatment outcomes in a myriad of other ocular conditions including cataract (e.g., AMD, glaucoma, keratoconus, corneal transplantation, low vision rehabilitation, and even in population-based studies).^{2–14} Moreover, several modified versions of the instruments have been put forward.^{15–22} The VF-14 has been adapted in several languages.^{4,7,10}

The VF-14 was initially developed and validated using the traditional method of psychometric test (the Classical Test Theory, CTT).¹ For several years, the summary scores of the VF-14 were predominantly used in ophthalmic research. However, lately the VF-14 has been tested with sophisticated psychometric assessment methods such as Rasch analysis.^{7,10,16,17,23} Rasch analysis is a probabilistic mathematical method that provides a

deeper insight into critical psychometric properties of a PRO instrument.^{24,25} The Rasch-based psychometric properties enable the assessment of an instrument's measurement quality against an established framework of stringent quality criteria.^{25,26} The other important advantage of Rasch analysis is that it provides interval-level estimates from ordinal instrument responses. Interval-level scores reduce noise with increased measurement precision; therefore, they allow increased statistical power to test the study hypotheses.²⁷ The direct implication of increased statistical power is that it may help to minimize the cost of clinical research by significantly reducing the sample size required.²⁸

Several research groups have used Rasch analysis to shorten the VF-14 instrument.^{7,16,17} More recently, Gothwal et al.¹⁶ put forward two shorter versions of the VF-14: the VF-11R and the VF-8R. The R stands for Rasch, given that Rasch analysis was used to justify the revisions and provide interval scoring. Among all the modified shorter versions, the VF-8R was reported to be the most responsive in measuring cataract surgical outcome.¹⁶

TABLE 1. Item Content for the English and the Chinese-Translated VF-14 Instruments

Item	English Version	Chinese-Translated Version
1	Reading small print, such as labels on medicine bottles, a telephone book, food labels	Reading small print, such as labels on medicine bottles, a telephone book, price list, watch
2	Reading a newspaper or a book	Reading a newspaper or a book
3	Reading a large-print book or large-print newspaper or numbers on a telephone	Reading large font, such as a large-print book or newspaper, numbers on a telephone or mobile phone, wall clock
4	Recognizing people when they are close to you	Recognizing familiar people when they are close to you
5	Seeing steps, stairs, or curbs	Seeing steps, stairs, or curbs
6	Reading traffic signs, street signs, or store signs	Reading signs, such as traffic signs, street signs, store signs, advertising board, or plate number
7	Doing fine handwork, such as sewing, knitting, crocheting, carpentry	Doing fine handwork, such as threading a needle, sewing, knitting, crocheting
8	Writing checks or filling out forms	Signing your name or filling out forms
9	Playing games, such as bingo, dominos, card games, mahjong	Playing games, such as card games, mahjong, chess
10	Taking part in sports, such as bowling, handball, tennis, golf	Taking part in sports, such as playing Ping-Pong or badminton, strolling, doing exercise, shadowboxing
11	Cooking	Cooking
12	Watching television	Watching television
13	Day driving	Day driving such as automobile, motorcycle, or nonmotorized vehicle
14	Night driving	Night driving such as automobile, motorcycle, or nonmotorized vehicle

Although, the VF-14 has been used in different cultural settings, only one study has reported using the instrument in a Chinese population.²⁹ However, the VF-14 has never been tested with Rasch analysis in Chinese settings.²⁹ Therefore, the aim of this study was to assess the VF-14 using Rasch analysis and test whether the native version and the previously proposed Rasch-validated shorter versions of the VF-14 could be adopted in a Chinese population. Additionally, the secondary aim was to determine whether it was necessary to create new versions of the VF-14 in China other than those proposed previously.

METHODS

Study participants were patients planned for cataract surgery at the Eye Hospital of Wenzhou Medical University, Wenzhou, China. The study participants completed the Chinese-translated VF-14 questionnaire by face-to-face interviews before undergoing cataract surgery. The exclusion criteria were patients with cognitive impairment and other ocular comorbidities affecting their vision. All participants provided written informed consent. The study was approved by the Review Board of Wenzhou Medical University, and it was carried out following the tenets of the Declaration of Helsinki for research involving humans.

The VF-14 Instrument

In brief, the VF-14 is a PRO instrument in which patients rate their difficulty in performing 14 vision-related daily living activities with their existing optical corrections.¹ Each item in the VF-14 is divided into two parts: part 1: "Do you have difficulty, even with glasses...?" Yes / No / Not applicable, where respondents can choose one of the answers that is relevant to them. If they choose "Yes," then they have to go on to the second part of the question; part 2: "If yes, how much difficulty do you currently have?" A little / A moderate amount / A great deal / Unable to do the activity. For this

study, the responses were scored on a five-point Likert scale from 0 ("No") to 4 ("Unable to do the activity") for all items. The response category "not applicable" was considered missing data.

For this study, the VF-14 was translated from English to Chinese (into Mandarin) independently by two medical doctors who are fluent in both languages. The two versions were reconciled by a panel of experts to produce a second draft. The draft was then translated back into English and compared with the original English version to identify any discrepancies between the two versions, which were then revised by the panel. The Chinese-translated VF-14 was then tested in 20 patients for its comprehension. Further revisions on the wording of each item were carried out on the basis of patients' feedback, and it was deemed necessary by the panel to match the Chinese socio-cultural norms to enhance item comprehension. The following is an example of necessary changes that were made—item 10: "Taking part in sports, such as bowling, handball, tennis, golf" was changed to "Taking part in sports, such as playing Ping-Pong or badminton, strolling, doing exercise, and shadowboxing." This was done because sports such as bowling, handball, tennis, and golf are not common in China and especially in the elderly Chinese population. Similarly, driving a car is not common in the elderly Chinese population; the wordings of the driving items (item 13 and item 14) were also changed. This exercise enabled us to culturally adapt the VF-14 in a Chinese socio-cultural scenario and at the same time maintain conceptual equivalence (i.e., activity limitations) for all the items between the English and the translated versions (Table 1). Even though, the VF-14 was translated into Mandarin, it was interviewer administered in local dialects (e.g., Wenzhounese) if the participants did not understand Mandarin.

Analysis Strategy

We performed Rasch analysis in three different stages. In Rasch stage I, the VF-14 was assessed in its full-length version. In Rasch stage II, we tested the validity, in a Chinese population,

TABLE 2. Demographic Characteristics of the Participants

Characteristics	Results
Median age, y (IQR)	71 (64-77; 40-92)
Sex, <i>n</i> (%)	
Female	265 (58)
Male	191 (42)
First eye surgery (%)	380 (83.3)
Second eye surgery (%)	76 (16.7)
Visual acuity; logMAR median, range (Snellen equivalent, range)	
Worse eye	1.20, 0.10 to 3.00 (20/138, 20/25 to NPL)
Better eye	0.64, -0.08 to 2.28 (20/59, 20/16 to HM)
Ocular comorbidity,* <i>n</i> (%)	210 (46.1)
Glaucoma, <i>n</i> (%)	14 (3.1)
AMD, <i>n</i> (%)	13 (2.9)
DR, <i>n</i> (%)	14 (3.1)
Pathologic myopia, <i>n</i> (%)	30 (6.6)
Corneal disorders, <i>n</i> (%)	20 (4.4)
Other, <i>n</i> (%)	149 (32.7)
Systemic comorbidity,† <i>n</i> (%)	312 (68.4)
Hypertension, <i>n</i> (%)	214 (46.9)
Diabetes, <i>n</i> (%)	87 (19.1)
Others, <i>n</i> (%)	166 (36.4)
Illiterate	222 (48.7)
Primary school	114 (25.0)
Junior middle school	73 (16.0)
Senior middle school	31 (6.8)
University	16 (3.5)

IQR, interquartile range; NPL, no perception of light; HM, hand motion.

* Includes diabetic retinopathy (DR), glaucoma, AMD, corneal disorders (e.g., corneal macula, corneal dystrophies), pathologic myopia, and other eye diseases (e.g., pterygium, vein occlusion, uveitis, epiretinal membrane).

† Percentages of comorbidities add more than the total sum, as some ocular and systemic conditions coexist.

of the shorter, modified scales put forward by Gothwal et al.,¹⁶ and at the same time we attempted to improve any deficiency identified in these scales (phase I analysis: Rasch stages I and II). In stage III, we developed a new, valid, shorter scale of the VF-14 from scratch (phase II analysis: Rasch stage III). The Chinese scales were developed following a Rasch-based iterative item removal criteria. For this, items were culled on the basis of missing data (missing data > 50%), fit statistics, and principal component analysis (PCA) of the residuals.^{24,30} Finally, we assessed correlations and agreement between the person scores obtained from the new shorter scales we developed in phase II and the scales put forward by Gothwal et al.¹⁶ (phase III analysis).

Rasch Analysis

The Winsteps program (Version 3.75.0, Winsteps, Beaverton, Oregon, USA) was used for Rasch analysis using the Andrich rating scale model. Rasch analysis is a mathematical probabilistic model that estimates a person's ability in relation to item difficulty expressed in log odds units (logits) on a single continuum scale. For this analysis, participants with higher ability and items of greater difficulty were located on the negative side of the continuum scale and vice versa. The following Rasch-based psychometric parameters were tested in sequence.

Category Threshold Order

The quality of an instrument is determined by the extent to which the response categories are used in an orderly fashion.³¹ Disordering of categories occurs when categories are underutilized, have unclear definitions, or the number of categories exceeds what the respondents can distinguish.³¹ Generally, the disordered categories are repaired by combining categories.³²

Measurement Precision

Measurement precision is one of the most important parameters of an instrument, indicating whether the instrument functions as a stand-alone measuring scale. It is estimated by person separation reliability (PSR) or index (PSI).³² Both PSR and PSI are measures of precision. A PSR of ≥ 0.8 (PSI ≥ 2.00) indicates that the instrument can distinguish the study population into three levels of disability (i.e., mild, moderate, or severe).³²

Unidimensionality

Unidimensionality is a fundamental requirement of a measurement scale. Unidimensionality refers to an instrument's ability to measure only one single underlying trait (e.g., activity limitation, symptoms, emotional well-being). An instrument that violates the conditions of unidimensionality does not meet the fundamental requirement of construct validity. Unidimensionality can be assessed loosely by examining the item fit statistics. The item fit statistics are expressed in mean square statistics. There are two types of fit statistics (infit and outfit), both measure how well the items fit the construct. A stringent criterion put forward by Pesudovs et al.²⁵ for fit statistics is between 0.7 and 1.3. However, a more lenient criterion between 0.5 and 1.5 is also considered useful for the measurement.^{30,33} The principal component analysis of the residuals (PCA) is a stringent test of dimensionality. In the PCA analysis, a level of 60% of the variance explained by the raw data is considered an indication of unidimensionality.^{24,25,34} Similarly, the first contrast in the residuals indicates whether there are other constructs within the variance that are not explained by the principal component, which suggests that a second construct is being measured. In the current study, we considered an eigenvalue > 2.0 as evidence of a second construct being measured by the instrument.²⁴

Targeting

Targeting refers to how well the difficulty of items matches the abilities of the study sample. The targeting is assessed by calculating the difference between the person and item means on the person-item map. A difference in means greater than 1 logit indicates notable mistargeting; an instrument that has perfect targeting would have a value of zero.^{24,35}

Differential Item Functioning (DIF)

Differential item functioning assesses whether items are responded to distinctly differently by population subgroups, stratified by the population's demographic characteristics. For this study, we assessed DIF of each item by age (≤ 50 years and > 50 years), sex, ocular comorbidity (present/absent), systemic comorbidity (present/absent) and better eye visual acuity ($\leq 6/60$ and > 6/60), and first or second cataract surgery. Small or absent DIF was defined as a difference in item measure of < 0.50 logits, minimal DIF as 0.50 to 1.0 logits, and notable DIF as > 1.0 logits. In this study, only DIFs defined as notable are reported.²⁴

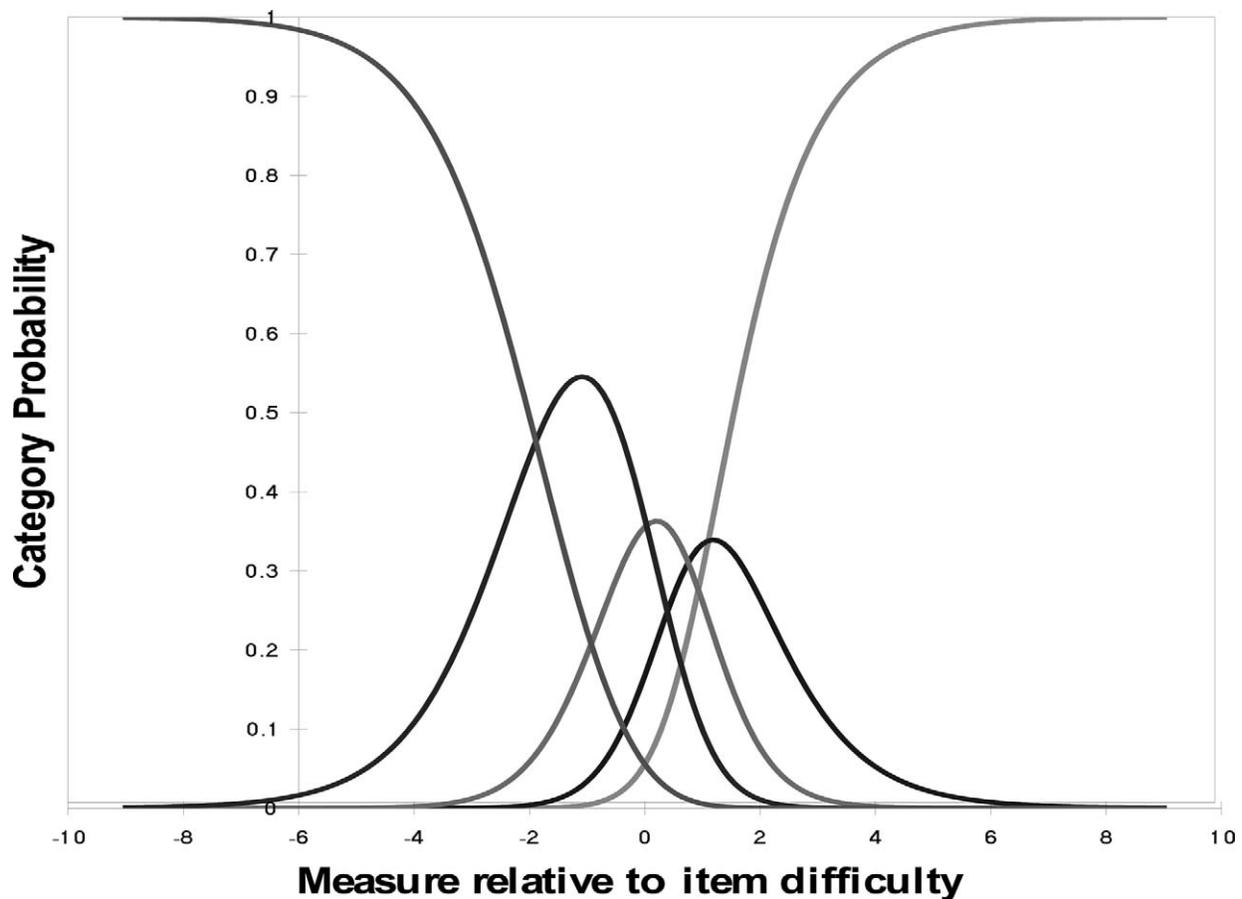


FIGURE 1. Category probability curves show the operation of the five-response categories. The curve at the *extreme left* represents “no difficulty,” and the curve at the *extreme right* represents “unable to do the activity.”

Correlation and Agreement

Correlation between the person scores obtained from Gothwal et al.¹⁶ proposed scales and the modified Chinese versions was tested with Spearman rank correlation coefficient. Bland-Altman 95% limit of agreement (LoA) plots were used to evaluate the agreement between scores obtained from the shorter scales put forward by Gothwal et al.¹⁶ and the new Chinese shorter scales developed in this study. A Bland-Altman 95% LoA plot is a method to assess the agreement between two methods of measurement or scales intended to measure the same construct.^{36,37} The measures of agreement between the different models are displayed graphically in plots and described with 95% confidence intervals of the mean difference.

Validity Assessment

Validity (criterion validity) of the VF-14 and its scales was also measured.²⁵ For this, we assessed the correlation between instrument scores and visual acuity in the better and worse eyes of the study participants. Spearman correlation coefficient was used for this purpose. The Spearman correlation coefficient of <0.2 was considered weak; from >0.2 to <0.8 , moderate; and >0.8 , strong.

Statistical Analysis

Descriptive data were analyzed using SPSS software (IBM SPSS Statistics for Windows, Version 22.0.0., IBM Corp., Armonk,

NY, USA). Mann-Whitney U test was used to test significance between different demographic characteristics. Spearman rank correlation was used if one datum or both data were not distributed normally. A P value of <0.05 was considered statistically significant. Bland-Altman LoA plots were produced using MedCalc (MedCalc Version 12.4.0.0; Acaciaaan 22, B-8400, Ostend, Belgium).

RESULTS

A total of 456 patients (58% female) planned for cataract surgery completed the VF-14 instrument. There was no significant difference in median visual acuities in better eye (Mann-Whitney U test, $P = 0.35$) and worse eye (Mann-Whitney U test, $P = 0.574$) between males and females. More than half of the participants had systemic comorbidities and slightly less than half had ocular comorbidities (Table 2).

Phase I Analysis (Rasch Analysis of the Original and the Existing Shorter Versions of the VF-14)

The Chinese-translated VF-14 and its two shorter valid versions (VF-11R and VF-8R) proposed by Gothwal et al.¹⁶ demonstrated comparable or better psychometric properties than in the Australian population (Table 2). There was no disordering of response categories, which signifies that all the response categories were evenly endorsed by the respondents (Fig. 1). Moreover, these three translated scales were better targeted in Chinese patients than in Australian patients with cataract

TABLE 3. Rasch-Based Metric Properties of the Scales Proposed by Gothwal et al.¹⁶ and Developed in the Current Study

Parameter	English Version (Gothwal et al. ¹⁶)			Current Study				
	VF-14	VF-11R*	VF-8R†	VF-14	13-Item VF-14‡	VF-11R*	10-Item VF-11R§	VF-8R†
Number of items	14	11	8	14	13	11	10	8
Measurement precision	2.45	2.46	2.29	2.51	2.42	2.58	2.33	2.09
Misfitting items	2	0	0	1	0	1	0	0
Mean person location	-1.86	-2.57	-1.97	-0.59	-0.83	-0.57	-0.86	-0.14
PCA, eigenvalue first contrast	2.3	1.6	1.6	2.0	2.0	2.0	2.0	1.8

* In this version (VF-11R), items 10, 13, and 14 were deleted.

† In this version (VF-8R), items 3, 4, 10, 11, 13, 14 were deleted.

‡ In this version (VF-14), item 7 was deleted.

§ In this version (VF-11R), items 7, 10, 13, and 14 were deleted.

(Table 3). The PCA analysis of the VF-14 and the VF-11R revealed that the eigenvalues of the first contrast were at the cutoff value of 2.0. All other contrasts had eigenvalues < 2.00. The PCA of the VF-14, the VF-11R, and the VF-8R explained 66.8%, 66.9%, and 66.4%, respectively, of the raw variance. These values are all above the cutoff value of 60%. In both the original VF-14 and the VF-11R scales, removing one grossly misfitting item (in VF-14: item 7, infit = 1.70 and outfit = 1.61; and in VF-11R: item 7, infit = 1.64 and outfit = 1.58) resulted in a 13-item VF-14 and a 10-item VF-11R. These two scales demonstrated good metric properties (Table 3). The VF-8R scale had good psychometric properties and did not need any modification (Table 3). None of the scales demonstrated notable DIF by age, sex, education, ocular comorbidity, systemic comorbidity, visual acuity, and first versus second eye surgery.

Phase II Analysis (Development of Chinese Shorter Versions of the VF-14)

We also opted to develop completely new shorter versions of the VF-14 from scratch. Items 9, 13, and 14 had missing data > 50%. These items were removed from further analysis. The Chinese VF-11R (hereafter abbreviated as the “VF-11RChin”) had good targeting and acceptable PCA and fit statistics, except item 7, which was misfitting (infit = 1.64, outfit = 1.58) (Table 4). Hence, item 7 was removed, which resulted in a 10-item VF-11RChin. This new scale demonstrated good Rasch-based metric properties (Table 4). A further shortening of this scale was not warranted as there was no basis for removing remaining items (all items had perfect fit statistics with no or minimum item redundancy). Therefore, a Chinese version of the VF-8R was not created. There was no notable DIF in these two scales by age, sex, education, ocular comorbidity, systemic comorbidity, visual acuity, and first versus second eye surgery.

TABLE 4. Shorter Versions of the VF-14 Developed in the Current Study

Parameter	Chinese Versions	
	VF-11RChin*	10-Item VF-11RChin†
Number of items	11	10
PSR/PSI	2.40	2.30
Misfitting items	1	0
Mean person location	-0.72	-1.04
PCA, eigenvalue first contrast	2.0	2.0

* In this version (VF-11RChin), items 9, 13, and 14 were deleted.

† In this version (10-Item VF-11RChin), items 7, 9, 13, and 14 were deleted.

Phase III Analysis (Testing Whether the Existing Shorter Versions of the VF-14 Could Be Applied in Chinese Population)

High correlations were observed between the 10-item VF-11RChin (developed in this study) and the scale proposed by Gothwal et al.¹⁶ (10-item VF-11R; $r = 0.996$, $P < 0.001$). In terms of agreement, there was a narrow 95% LoA (mean difference, -0.18; 95% LoA, -0.47 to 0.11), which implied good agreement and interchangeability between these two scales (Fig. 2).

Validity Assessment

The construct validity of the 13-item VF-14, 10-item VF-11R and VF-8R was tested by exploring their correlation with both the better eye and the worse eye visual acuities. The correlations were moderate but statistically significant for both eyes (Table 5). However, the correlations were stronger between the visual function scores and better eye visual acuity, probably because better eye visual acuity is equivalent to binocular visual acuity (Table 5). This demonstrated the construct validity of the scales.

Ready-to-Use Scoring Spreadsheets

We have also developed ready-to-use Microsoft Excel spreadsheets to ease the use of the three scales of the VF-14 in a Chinese population (see Supplementary Material). The spreadsheets can be used to estimate person scores in logits without having to do Rasch analysis in patients with similar demographics, as in this study. Each spreadsheet has three sheets labeled as “rawdata,” “raschscore,” and “raw to rasch conversion.” Users only need to register respondents’ VF responses to items in numerical label (i.e., 0-4) in the “rawdata” sheet to obtain the corresponding Rasch scores in the “raw to rasch conversion” sheet. The translated VF-14 instrument in Mandarin is available as Supplementary Material or can be obtained from the corresponding author.

DISCUSSION

In the backdrop of several versions of the VF-14 questionnaire being available,¹⁶⁻¹⁸ it was a dilemma for us to choose the version that would best suit the Chinese cataract population. It was not practical to adopt all of them in our setting. Moreover, many of these versions were developed using traditional validation methods.¹⁶ Therefore, we set out to test whether the original version and its two shorter versions (the VF-11R and the VF-8R) were applicable in a Chinese cataract population, using Rasch analysis. We chose the VF-11R and

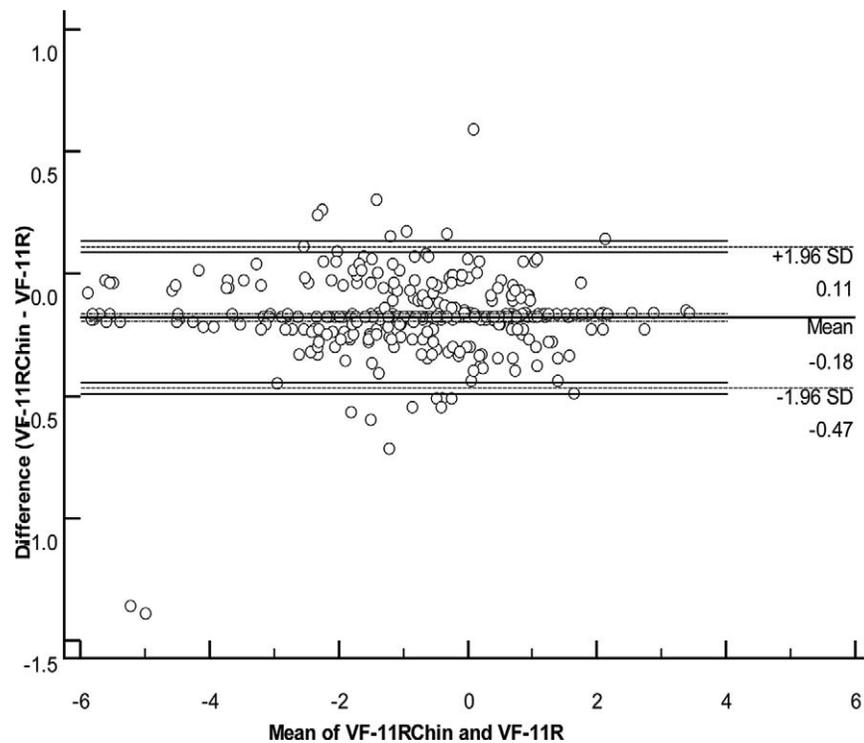


FIGURE 2. Bland-Altman plot of agreement between the Gothwal et al.¹⁶ 10-item VF-11R and the 10-item VF-11RChin.

the VF-8R because they were the best among all the existing shorter versions of the VF-14 in terms of Rasch-based metric properties, with excellent precision and unidimensionality as well as being highly responsive in measuring cataract outcomes.^{16,26} The main objective of this study, therefore, was to translate, culturally adapt, and revalidate these three scales (the original VF-14, the VF-11R, and the VF-8R) in a Chinese population. The reason for this approach was to set out a clear understanding of whether it was necessary to develop new versions of the VF-14 for the Chinese population while valid and appropriate performing versions are available in English. We believe that developing slightly different versions of the same instrument for every population and in every language is confusing for everyone who wants to use a particular instrument with a given brand name. Further, the use of different versions of the same instrument may limit direct comparison between studies conducted in different countries and populations. We believe that the evidence provided in this study challenges and may discourage the current culture of developing new versions and versions of the same PRO instrument without considering the applicability of existing versions.

TABLE 5. Relationship Between the Different Versions of the VF-14 and Visual Acuity

	Better Eye VA, Spearman ρ^*	Worse Eye VA, Spearman ρ^*
13-Item VF-14	0.486	0.287
10-Item VF-11R	0.478	0.284
VF-8R	0.455	0.285

VA, visual acuity.

* Statistically significant at <0.001 (2-tailed).

This study demonstrated that the VF-14 and its two shorter scales proposed by Gothwal et al.¹⁶ can be adapted to assess cataract surgical outcomes in a Chinese population without having to develop new versions. We have had similar findings when we tried to adapt the existing valid versions of the National Eye Institute Visual Function Questionnaire (NEI VFQ) in Chinese settings.³⁸ Both the VF-14 and the VF-11R had a grossly misfitting item (item 7: doing handwork). The subsequent removal of item 7 from both scales resulted in unidimensional and psychometrically robust scales (i.e., the 13-item VF-14 and the 10-item VF-11R) with no misfitting items (Table 3). Similar findings and item deletion were carried out in the Spanish and German versions of the VF-14.^{7,10,23} And, the VF-8R did not need any modification in our study, as the scale demonstrated appropriate Rasch-based metric properties. Therefore, these three scales were perfectly poised to be used in Chinese settings.

Moreover, when compared with the Gothwal et al. study,¹⁶ these three scales demonstrated superior metric properties and excellent targeting in the Chinese cataract population (Table 3).¹⁶ Of interest, the two-level question format did not dilute the targeting of the scales in this study, unlike the findings of previous studies carried out in Western countries.^{10,23,39} The improved metric properties including targeting of these scales to our study population could also be attributed to the careful translation and cultural adaptation used in this study to reflect the Chinese socio-cultural practice. More important, this study indicates that the items of the VF-14 are still relevant for developing nations such as China, where there is a lower literacy rate and socio-economic index than in Western countries.^{17,23,40}

The VF-14 was translated in Mandarin, and then necessary modifications were made for cultural adaptation to match Chinese socio-cultural practice in general. The translation in Mandarin was a conscious decision because it is the official language and the official Chinese script, and is widely spoken

in China. Notably, the majority of other Chinese dialects (e.g., Wenzhounese and Cantonese) are spoken differently, but they have the same script as that of Mandarin. Moreover, over 70% of our study population spoke and understood Mandarin. Therefore, in the majority of cases, administration of the VF-14 was not an issue. However, patients who speak different dialects (e.g., Wenzhounese, Cantonese) are not uncommon at the Eye Hospital of Wenzhou. In order to minimize the influence of different dialects on the validity of the VF-14 in this study, Mandarin was used priority to other dialects when investigating. Utmost caution was practiced when administering the instrument to patients who spoke different dialects. Basically, our trained staff communicated in the interviewee's dialect while administering the VF-14 if Mandarin could not be understood well. We also made sure that the verbal translation in other dialects was in accordance with the original meaning of the instrument both in meaning and phraseology. Moreover, face-to-face administration of the VF-14 and the strict principle of translation during administration have minimized the risk of any deviation or misunderstanding between different dialects of the VF-14 items. The question of whether the VF-14 is a valid instrument in Chinese people living outside China is debatable. This is because of the vast difference in economic, cultural, health status, and lifestyle among Chinese decedents living outside China and those in mainland China. Further studies are warranted for cultural adaptation and to test validity in those populations.

A PRO instrument developed for one population may not be directly relevant in another population that differs vastly in terms of sociodemographics. Therefore, there is a great value in cross-cultural validation in creating a more relevant and responsive instrument. However, the current practice of developing completely new or slightly different versions of the existing instruments measuring the same construct (e.g., activity limitations, symptoms) is less appealing as clinicians and researchers grapple to choose an appropriate instrument from the plethora of existing instruments. At present, there are over 125 different paper-pencil-based PRO instruments in ophthalmology and optometry alone. Taking into account different languages and shorter versions, the number of existing PRO instruments swells roughly to 200. The current trend of adding yet another new instrument within this huge pool of instruments is counterproductive in advancing the field of PRO measurement in optometry and ophthalmology.

The way forward is to develop technologically advanced PROs in the form of item banks implemented via computer-adaptive testing (CAT) system.⁴¹ An item bank incorporated with the CAT system is more flexible and allows an easy evolution of items to match the changing time. A similar initiative; the Patient-Reported Outcomes Measurement Information System (PROMIS, <http://www.nihpromis.org>; provided in the public domain by the National Institutes of Health, Bethesda, MD, USA) has revolutionized PRO outcome measurement in various diseases and populations.⁴² Our research group is currently developing such technologically advanced instruments for all eye diseases (the Eye-tem Bank).⁴³⁻⁴⁵ The Eye-tem Bank is a National Health and Medical Research Council (NHMRC)-funded project that aims to develop technologically advanced patient-reported outcomes in the form of item banks implemented via CAT system for all eye diseases across all populations.⁴³⁻⁴⁵ We encourage researchers around the world to join us in the development of the Eye-tem Bank to make it relevant across all populations worldwide.

The VF-14 instrument has been trialed and tested in ophthalmology many times before; however, such studies were primarily carried out in developed countries.^{8,10,14,16,21} To the best of our knowledge, the VF-14 instrument has neither been validated nor has it undergone rigorous psychometric

testing in a Chinese population before. Of interest, several studies carried out in countries with higher socio-economic status have shown that the VF-14 is less sensitive or a poorly targeted to measure cataract outcomes (e.g., in Australia or Germany) when compared with other cataract-specific instruments.^{2,10,16} The reason for this might be because of the way items in the VF-14 are worded or that the item contents are outdated.³⁹ However, this study shows that the VF-14 is psychometrically robust and perfectly targeted to the study population. These findings signify that the instrument is still very relevant and significant to Chinese people with cataract. Taking this in view, this validated VF-14 has the potential to be widely used as a clinical and research outcome measure in China, where the rate of cataract-related blindness is increasing exponentially. Therefore, we expect this article will be a significant contribution to the literature.

In conclusion, this study has provided evidence that it is not always necessary to develop a different version of a PRO instrument, as the existing versions may turn out to be relevant. Therefore, we suggest researchers adapt the existing valid Rasch-analyzed versions rather than embarking on a path of developing new or different versions of the existing instrument blindly.

Acknowledgments

Supported in part by the Foundation of Wenzhou City Science & Technology Bureau (Grant Y20140159); the innovative project of the School of Optometry and Ophthalmology and Eye Hospital, Wenzhou Medical University (Grant YNCX201301), and the National Science and Technology Ministry (2012BAI08B04). The authors alone are responsible for the content and writing of the paper.

Disclosure: **J. Khadka**, None; **J. Huang**, None; **K. Mollazadegan**, None; **R. Gao**, None; **H. Chen**, None; **S. Zhang**, None; **Q. Wang**, None; **K. Pesudovs**, None

References

- Steinberg EP, Tielsch JM, Schein OD, et al. The VF-14. An index of functional impairment in patients with cataract. *Arch Ophthalmol*. 1994;112:630-638.
- McAlinden C, Gothwal VK, Khadka J, Wright TA, Lamoureux EL, Pesudovs K. A head-to-head comparison of 16 cataract surgery outcome questionnaires. *Ophthalmology*. 2011;118:2374-2381.
- Linder M, Chang TS, Scott IU, et al. Validity of the visual function index (VF-14) in patients with retinal disease. *Arch Ophthalmol*. 1999;117:1611-1616.
- Maharajah KR, Tet CM, Yaacob A, Tajudin IS, Foster PJ. Modified Bahasa Malaysia version of VF-14 questionnaire: assessing the impact of glaucoma in rural area of Malaysia. *Clin Experiment Ophthalmol*. 2008;36:222-231.
- Riusala A, Sarna S, Immonen I. Visual function index (VF-14) in exudative age-related macular degeneration of long duration. *Am J Ophthalmol*. 2003;135:206-212.
- Sawada H, Fukuchi T, Abe H. Evaluation of the relationship between quality of vision and the visual function index in Japanese glaucoma patients. *Graefes Arch Clin Exp Ophthalmol*. 2011;249:1721-1727.
- Valderas JM, Alonso J, Prieto L, Espallargues M, Castells X. 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*. 2004;13:35-44.
- Ziakas N, Kanonidou E, Mikropoulos D, Georgiadis N. The impact of penetrating keratoplasty in patients with keratoconus using the VF-14 questionnaire. *Int Ophthalmol*. 2011;31:89-92.

9. Boisjoly H, Gresset J, Charest M, et al. The VF-14 index of visual function in recipients of a corneal graft: a 2-year follow-up study. *Am J Ophthalmol*. 2002;134:166-171.
10. Chiang PP, Fenwick E, Marella M, Finger R, Lamoureux E. Validation and reliability of the VF-14 questionnaire in a German population. *Invest Ophthalmol Vis Sci*. 2011;52:8919-8926.
11. Khor WB, Wei RH, Lim L, Chan CM, Tan DT. Keratoconus in Asians: demographics, clinical characteristics and visual function in a hospital-based population. *Clin Experiment Ophthalmol*. 2011;39:299-307.
12. Saw SM, Foster PJ, Gazzard G, Seah S. Causes of blindness, low vision, and questionnaire-assessed poor visual function in Singaporean Chinese adults: the Tanjong Pagar Survey. *Ophthalmology*. 2004;111:1161-1168.
13. Alonso J, Espallargues M, Andersen TF, et al. International applicability of the VF-14: an index of visual function in patients with cataracts. *Ophthalmology*. 1997;104:799-807.
14. Bilbao A, Quintana JM, Escobar A, et al. Responsiveness and clinically important differences for the VF-14 index, SF-36, and visual acuity in patients undergoing cataract surgery. *Ophthalmology*. 2009;116:418-424. e411.
15. Cillino S, Casuccio A, Di Pace F, et al. One-year outcomes with new-generation multifocal intraocular lenses. *Ophthalmology*. 2008;115:1508-1516.
16. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Measuring outcomes of cataract surgery using the Visual Function Index-14. *J Cataract Refract Surg*. 2010;36:1181-1188.
17. Lamoureux EL, Pesudovs K, Thumboo J, Saw SM, Wong TY. An evaluation of the reliability and validity of the visual functioning questionnaire (VF-11) using Rasch analysis in an Asian population. *Invest Ophthalmol Vis Sci*. 2009;50:2607-2613.
18. Uusitalo RJ, Brans T, Pessi T, Tarkkanen A. Evaluating cataract surgery gains by assessing patients' quality of life using the VF-7. *J Cataract Refract Surg*. 1999;25:989-994.
19. Pager CK. Assessment of visual satisfaction and function after cataract surgery. *J Cataract Refract Surg*. 2004;30:2510-2516.
20. Moghimi S, Esfahani MR, Maghsoudipour M. Visual function after implantation of aniridia intraocular lens for traumatic aniridia in vitrectomized eye. *Eur J Ophthalmol*. 2007;17:660-665.
21. Friedman DS, Tielsch JM, Vitale S, Bass EB, Schein OD, Steinberg EP. VF-14 item specific responses in patients undergoing first eye cataract surgery: can the length of the VF-14 be reduced? *Br J Ophthalmol*. 2002;86:885-891.
22. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Med Care*. 2004;42:17-24.
23. Las Hayas C, Bilbao A, Quintana JM, Garcia S, Lafuente I. A comparison of standard scoring versus Rasch scoring of the visual function index-14 in patients with cataracts. *Invest Ophthalmol Vis Sci*. 2011;52:4800-4807.
24. Khadka J, Pesudovs K, McAlinden C, Vogel M, Kernt M, Hirneiss C. Reengineering the glaucoma quality of life-15 questionnaire with Rasch analysis. *Invest Ophthalmol Vis Sci*. 2011;52:6971-6977.
25. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci*. 2007;84:663-674.
26. Khadka J, McAlinden C, Pesudovs K. Quality assessment of ophthalmic questionnaires: review and recommendations. *Optom Vis Sci*. 2013;90:720-744.
27. Khadka J, McAlinden C, Pesudovs K. Cognitive trajectories after postoperative delirium. *N Engl J Med*. 2012;367:1164.
28. Chakravarty EF, Bjonner JB, Fries JF. Improving patient reported outcomes using item response theory and computerized adaptive testing. *J Rheumatol*. 2007;34:1426-1431.
29. Luo M, Ji J, Zhao C, Fan X. Clinical study of Acrysof IQ aspheric intraocular lenses. *Clin Experiment Ophthalmol*. 2010;38:358-362.
30. Pesudovs K, Garamendi E, Elliott DB. The Quality of Life Impact of Refractive Correction (QIRC) Questionnaire: development and validation. *Optom Vis Sci*. 2004;81:769-777.
31. Khadka J, Gothwal VK, McAlinden C, Lamoureux EL, Pesudovs K. The importance of rating scales in measuring patient-reported outcomes. *Health Qual Life Outcomes*. 2012;10:80.
32. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 2nd ed. London: Lawrence Erlbaum Associates; 2007.
33. Khadka J, Ryan B, Margrain TH, Court H, Woodhouse JM. Development of the 25-item Cardiff Visual Ability Questionnaire for Children (CVAQC). *Br J Ophthalmol*. 2010;94:730-735.
34. Linacre JM. *A User's Guide to WINSTEPS*. Chicago, IL: Winsteps; 2012. <http://www.winsteps.com>. Accessed October 10, 2013.
35. McAlinden C, Khadka J, de Freitas Santos Paranhos J, Schor P, Pesudovs K. Psychometric properties of the NEI-RQL-42 questionnaire in keratoconus. *Invest Ophthalmol Vis Sci*. 2012;53:7370-7374.
36. McAlinden C, Khadka J, Pesudovs K. Statistical methods for conducting agreement (comparison of clinical tests) and precision (repeatability or reproducibility) studies in optometry and ophthalmology. *Ophthalmic Physiol Opt*. 2011;31:330-338.
37. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307-310.
38. Mollazadegan K, Huang J, Khadka J, et al. Cross-cultural validation of the National Eye Institute Visual Function Questionnaire. *J Cataract Refract Surg*. 2014;40:774-784.
39. Khadka J, McAlinden C, Gothwal VK, Lamoureux EL, Pesudovs K. The importance of rating scale design in the measurement of patient-reported outcomes using questionnaires or item banks. *Invest Ophthalmol Vis Sci*. 2012;53:4042-4054.
40. Quintana JM, Garcia S, Aguirre U, et al. Relationship of sociodemographic variables with outcomes after cataract surgery. *Eye (Lond)*. 2013;27:698-707.
41. Pesudovs K. Item banking: a generational change in patient-reported outcome measurement. *Optom Vis Sci*. 2010;87:285-293.
42. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010;63:1179-1194.
43. Fenwick EK, Pesudovs K, Khadka J, et al. The impact of diabetic retinopathy on quality of life: qualitative findings from an item bank development project. *Qual Life Res*. 2012;21:1771-1782.
44. Fenwick EK, Pesudovs K, Khadka J, Rees G, Wong TY, Lamoureux EL. Evaluation of item candidates for a diabetic retinopathy quality of life item bank. *Qual Life Res*. 2013;22:1851-1858.
45. Khadka J, McAlinden C, Craig JE, Fenwick EK, Lamoureux EL, Pesudovs K. Identifying content for the glaucoma-specific item bank to measure quality-of-life parameters [published online ahead of print April 2, 2013]. *J Glaucoma*. doi:10.1097/IJG.0b013e318287ac11.