Validating the Sumi Quality of Life Questionnaire With Rasch Analysis

Hiroyo Hirasawa, Hiroshi Murata, Chihiro Mayama, and Ryo Asaoka

Department of Ophthalmology, University of Tokyo Graduate School of Medicine, Tokyo, Japan

Correspondence: Ryo Asaoka, Department of Ophthalmology, University of Tokyo, Graduate School of Medicine, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655, Japan; rasaoka-tky@umin.ac.jp
Submitted: March 19, 2014
Accepted: June 19, 2014
Citation: Hirasawa H, Murata H, Mayama C, Asaoka R. Validating the Sumi Quality of Life Questionnaire with Rasch analysis. Investig Ophthalmol Vis Sci. 2014;55:5776–5782. DOI: 10.1167/iovs.14-14390

PURPOSE. To validate the Sumi Visual Disability Questionnaire (Sumi VDQ) to estimate visual disability in glaucoma patients using Rasch analysis.

METHODS. A total of 162 glaucoma patients underwent visual field (VF) testing in both eyes (Humphrey 24-2 Swedish Interactive Threshold Algorithm [SITA] standard program). The binocular VF was then calculated using the integrated VF (IVF) method. Visual disability was scored using the Sumi VDQ, which was originally written in Japanese and contains 50 items (questions). Response scale analysis, targeting, and infit statistics associated with Rasch analysis were evaluated. Unidimensionality was analyzed using principal component analysis (PCA). In addition, the correlation between the person parameter obtained with Rasch analysis and the mean of total deviation values (mTD) in the IVF was compared with the correlation between the arithmetic sum of visual disability score and mTD of the IVF.

RESULTS. All 30 items in the Sumi VDQ showed productive infit values (0.61–1.46). The person parameters distributed between −4.50 and 3.62, while the item difficulty targeting parameters distributed between −0.88 and 2.06. None of the PCA components had eigenvalues whose lower limit of 95% confidence interval (CI) exceeded 2 (0–1.5). There was a significant relationship between person parameter and mTD of IVF \( r = -0.78, \ P < 0.001 \), which was significantly stronger (Meng-Rosenthal-Rubin method, \( P = 0.002 \)) than that between arithmetic sum of visual disability score and mTD of IVF \( r = -0.61, \ P < 0.001 \).

CONCLUSIONS. The Sumi VDQ has constructive psychometric properties. In particular, the Rasch analysis–derived person parameter appears to be clinically more meaningful than the arithmetic sum of visual disability score.

Keywords: quality of vision, Rasch analysis, questionnaire, glaucoma

Q uality of vision can be defined as a person’s satisfaction with his or her visual ability and how the person’s vision impacts his or her daily life. Glaucoma, the second leading cause of blindness in the world, causes visual field (VF) damage and reduced visual acuity (VA), which impact glaucoma patients’ perception of his or her vision damage from what can be inferred from the grayscale output on a VF chart obtained by standard automated perimetry. It is important to estimate glaucoma patients’ visual disability using appropriate questionnaires so that clinicians can appreciate how the disease impacts patients’ daily lives. Indeed, many questionnaires have been proposed to evaluate the visual disability of glaucoma patients.

There are two main methods for testing the psychometric properties of quality of life instruments, including classical test theory (CTT) and item response theory (IRT). Despite the popularity of CTT in the development of ophthalmological questionnaires, its limitations are widely acknowledged; in particular, it is unable to consider important aspects of the questionnaire measurement such as item difficulty, item discrimination, and ordering of response categories. Furthermore, Cronbach’s \( \alpha \), which is frequently used in CTT to investigate the internal consistency of a questionnaire, is artificially inflated with a greater number of items in the test. Items can be analyzed individually with respect to the amount of information they provide about the latent trait using IRT (scores from multiple items are simply added together in CTT). Rasch analysis is a special case of IRT, whereby items and responders can be scaled according to the series of responses made. Rasch analysis places items and persons on a linear scale and provides an “infit” statistic to indicate how well different items describe a group of subjects and how well individual subjects fit the group. These favorable attributes have made Rasch analysis a popular method for testing instrument validity and applicability.

The Sumi Visual Disability Questionnaire (VDQ) was developed in 2000 after revision of an earlier questionnaire; it includes detailed questions on a variety of tasks in daily life. The questionnaire, written in Japanese, also includes one item regarding the difficulty in reading vertically, since this is the traditional way to read and write sentences in many East Asian countries, including Japan. Using this questionnaire, Sumi et al. reported a close relationship between retinal VF sensitivity in the inferior hemifield within 5° of fixation and visual disability. More recently Murata et al. revealed the importance of peripheral VF areas for different daily tasks. Similar to other questionnaires, the Sumi VDQ was developed and validated using CTT and, until now, has not been validated using Rasch analysis. Thus, the aim of this study was to explore the psychometric properties of the Sumi VDQ using Rasch analysis. Finally, the relationship between the Sumi VDQ’s visual...
Validating the Sumi Quality of Life Questionnaire

Table 1. The Sumi Visual Disability Questionnaire

<table>
<thead>
<tr>
<th>Questions Included in the Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letters</td>
</tr>
<tr>
<td>1. Can you read the headlines of a newspaper? (Yes/With difficulty/No)</td>
</tr>
<tr>
<td>2. Can you read small print in a newspaper? (Yes/With difficulty/No)</td>
</tr>
<tr>
<td>3. Can you read words in a dictionary? (Yes/With difficulty/No)</td>
</tr>
<tr>
<td>4. Can you see the numbers in a telephone directory? (Yes/With difficulty/No)</td>
</tr>
<tr>
<td>5. Can you make out a fare table for trains and subways? (Yes/With difficulty/No)</td>
</tr>
<tr>
<td>Sentences</td>
</tr>
<tr>
<td>6. Do you have difficulty reading and writing? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>7. When you write sentences in vertical lines, does it lean to either direction? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>8. When you read, can you find the next line easily? (Yes/With difficulty/No)</td>
</tr>
<tr>
<td>Walking</td>
</tr>
<tr>
<td>9. Do you have difficulty walking because of your visual problems? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>10. Can you take a walk by yourself? (Yes/With difficulty/No)</td>
</tr>
<tr>
<td>11. Do you misjudge traffic signals? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>12. Do you bump into people or objects while walking? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>13. Do you stumble on the stairs? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>14. Do you fail to notice changes in the ground? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>15. Do you fail to recognize your friends until they talk to you? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>16. Do you fail to see people or cars approaching you from the side? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>Going out</td>
</tr>
<tr>
<td>17. Do you have difficulty going out because of your visual problems? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>18. Do you need somebody to accompany you to go to new places? (No/Preferably/Yes)</td>
</tr>
<tr>
<td>19. Can you get a cab by yourself? (Yes/With difficulty/No)</td>
</tr>
<tr>
<td>20. Do you have difficulty traveling by train? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>21. Do you feel uneasy going out at night because of your visual problems? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>Dining</td>
</tr>
<tr>
<td>22. Do you have difficulty dining because of your visual problems? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>23. Do you drop food while dining because of your visual problems? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>24. Do you spill tea while pouring into a cup? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>25. Do you have difficulty using chopsticks? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>Dressing</td>
</tr>
<tr>
<td>26. Do you ever button up clothing in the wrong order? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>27. Can you see your face clearly in the mirror? (Yes/With difficulty/No)</td>
</tr>
<tr>
<td>Miscellaneous</td>
</tr>
<tr>
<td>28. Can you recognize people’s faces on TV? (Yes/With difficulty/No)</td>
</tr>
<tr>
<td>29. Do you have difficulty finding objects dropped on the floor? (No/Occasionally/Frequently)</td>
</tr>
<tr>
<td>30. Do you have difficulty dialing the telephone? (No/Occasionally/Frequently)</td>
</tr>
</tbody>
</table>

Questions were originally written in Japanese.

disability score and traditional measurements of vision, namely VF sensitivity and visual acuity (VA), was investigated.

Methods

In 162 glaucoma patients, an interview on the perception of visual disability was performed by a single investigator (HH), who was not involved in the clinical examination and treatment of the patient’s glaucoma, using the visual disability questionnaire developed by Sumi et al. (described below). Interviews were conducted at the outpatient clinic of the University of Tokyo Hospital (Tokyo, Japan). Study inclusion criteria were as follows: (1) Glaucoma was the only disease causing VF damage and/or VA impairment; (2) patients had no physical impairments; (3) patients were followed for at least 6 months during which intraocular pressure and VF damage were stable; (4) the VF was evaluated using the Humphrey Field Analyzer (HFA; Carl Zeiss, Dublin, CA, USA) 30-2 Swedish Interactive Threshold Algorithm (SITA) standard program with reliable results (fixation losses < 25%; false-positive error < 15%; false-negative error rate was not used following previously published results); and (5) patients had a glaucomatous VF defect in at least one eye. Glaucomatous VF damage was defined as the presence of at least one of the following criteria (according to the method of Anderson and Patella): a pattern deviation probability plot showing a cluster of three or more points with a probability of less than 5% and at least one point with a probability less than 1% in an expected hemifield; a pattern standard deviation with a probability of less than 5%; or a glaucoma hemifield test result outside normal limits.

Written informed consent was gained from all patients. Study approval was obtained from the Ethics Board of the institute, and the tenets of the Declaration of Helsinki were followed.

Sumi Visual Disability Questionnaire

Visual disability was assessed using the method initially reported by Sumi et al. The questionnaire, originally written in Japanese, contains 30 items regarding seven tasks: legibility of letters (“letters”), legibility of sentences (“sentences”), walking, using public transportation (“going out”), dining, dressing, and additional miscellaneous activities (“miscellaneous”) (see Table 1; note that questions have been translated into English for this article). The Sumi VDQ also includes one item (question 7) regarding the difficulty in reading vertically, since this is the traditional way to read/write sentences in Japanese. All of the items are scored on a three-category difficulty scale, as follows: 2 = greatly disabled, 1 = slightly disabled, 0 = not disabled. Within 3 months of undertaking the Sumi VDQ, the patient’s VF was tested in both eyes.

Classical Test Theory Analysis

Internal consistency was calculated using Cronbach’s α statistic. Internal consistency represents the extent to which all the items in a test measure the same concept or construct and hence the interrelatedness of items within the test. The coefficient ranges in value from 0 to 1; the higher the score, the more reliable is the generated scale. A score of 0.7 or more is generally deemed acceptable.

Rasch Analysis

Rasch analysis is a probabilistic mathematical model that estimates item difficulty, person ability, and threshold for each
response category on a single continuum logit scale (log-odds units). This analysis enables persons and items to be positioned on a linear scale according to estimated item calibration values and persons’ visual disability.\textsuperscript{31,55} With respect to the Sumi VDQ, a positive item logit score indicates a more difficult item while a positive person logit score indicates lower quality of vision (QoV). Conversely, a negative item logit score indicates an easier item while a negative person logit score indicates higher visual disability. Rasch analysis was used to investigate the following assessments: response scale analysis, targeting, infit statistics, unidimensionality, person separation index, and differential item functioning (DIF), following a previous review.\textsuperscript{45}

Response scale analysis investigates whether response categories have distinct meaning (ordered thresholds in the category probability curves) and whether each category had equal probability to be endorsed by the participants (items evenly spaced). The category threshold is the crossover point between response categories and indicates the point at which the likelihood of choosing either response category is the same. All the items of the Sumi VDQ were scored on a three-category response scale of increasing difficulty and have two thresholds. Disorderling of the threshold can occur due to reasons such as the presence of too many categories or when the labeling of categories is potentially confusing.\textsuperscript{50}

Item targeting is a person–item map that provides a visual observation of the relative position of item difficulty to a person’s ability. Targeting refers to how well item difficulty matches the participant’s ability, and, for a well-targeted instrument, mean item difficulty is usually set at zero; the greater the difference of item and participant parameters, the poorer the targeting.\textsuperscript{25}

The infit statistic was measured as the mean square standardized residuals (MNSQ). Item infit < 0.7 indicates redundancy, and values higher than 1.3 indicate a high level of noise in the responses suggesting misfitting\textsuperscript{45}; however, values between 0.5 and 1.5 can still be considered productive.\textsuperscript{51}

Unidimensionality was assessed using the 95% confidence interval (CI) of the residuals resulting from principal component analysis (PCA) by carrying out fast and robust bootstrapping\textsuperscript{52} (10,000 iterations). Unidimensionality indicates that a score produced by a measure represents a single concept\textsuperscript{51} while multidimensionality indicates that there is evidence of an additional component being captured by the item.\textsuperscript{50,55} In the PCA, an eigenvalue > 2.00 U is suggestive of a second construct being measured, indicating a multidimensional instrument.

In Rasch analysis, person separation index is calculated as the ratio of true measure variance to observed measure variance. A person separation index less than 0.5 implies that the differences between persons are mainly due to measurement error.\textsuperscript{54} and previous papers have recommended a minimum value of 0.8.\textsuperscript{45,55,56} In addition, DIF was investigated using sex and age (younger than 60 years and 60 years or older). Differential item functioning refers to a measurement bias, observed when members belonging different groups with the same latent trait or ability have a different probability of giving a response on a questionnaire. In the current study, the significance of DIF was tested using the relative change in the $\beta$ coefficient logistic regression DIF analysis using IRT \textsuperscript{0} estimates as the conditioning variable.\textsuperscript{57}

For evaluating whether the Sumi instrument is consistent with the established measurements of visual function, we calculated Pearson's correlation coefficient between the person parameter (logits) and external clinical measures; specifically, the mean of all total deviation (mTD) values in the integrated visual field (IVF),\textsuperscript{50} in which a binocular VF was calculated for each patient by merging a patient's monocular

<table>
<thead>
<tr>
<th>Table 2. Demographic Data of the 162 Glaucoma Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Demographics, 162 Glaucoma Patients</td>
</tr>
<tr>
<td>Age, y</td>
</tr>
<tr>
<td>MD of better eye, dB</td>
</tr>
<tr>
<td>MD of worse eye, dB</td>
</tr>
<tr>
<td>VA of better eye, logMAR</td>
</tr>
<tr>
<td>VA of worse eye, logMAR</td>
</tr>
</tbody>
</table>

The data are expressed as the mean ± standard deviation range. MD, mean deviation; dB, decibels.

HFA VFs using the “best sensitivity” method and VAs of persons’ better and worse eyes. For comparison, we also calculated Pearson's correlation coefficient between the arithmetic sum of the visual disability score and these external clinical measures.

Statistical Analysis

All analyses were performed using the statistical programming language R (version 2.15.3; Foundation for Statistical Computing, Vienna, Austria). The R eRM package was used to carry out Rasch analysis, the stats package to carry out PCA, the FRB package to calculate the 95% CI of the PCA residuals, and the dft package to calculate the DIF. Comparison of the correlation coefficient between overlapping groups was carried out using the Meng-Rosenthal-Rubin method.\textsuperscript{59}

RESULTS

One hundred sixty-two glaucomatous patients (86 males and 76 females) were enrolled in this study, which included 85 patients with primary open-angle glaucoma, 70 patients with normal-tension glaucoma, 4 patients with primary angle-closure glaucoma, and 3 patients with secondary open-angle glaucoma. Characteristics of the study subjects are summarized in Table 2. In the current study, Cronbach’s $\alpha$ was 0.96.

For the response scale analysis, the category response thresholds for each of the 30 items were investigated. There were ordered thresholds between all response categories indicating that each category had distinct meaning. The person–item map is illustrated in Figure 1, and parameters of each item are shown in Table 3. The person parameters distributed between $−4.50$ and $3.62$. The means of item and person parameters were $0.17$ and $−2.0$ logits, and a difference of $2.17$ logits was observed. In the analysis of item fit, all 30 items showed productive infit values (range: $0.55$–$1.44$; see Table 3): $55.2\%$ of the variance of the amount of raw variance was explained by the measure. The eigenvalue of the PCA components of the residuals varied from $0.25$ to $2.7$; however, the lower limit of the $95\%$ CI never exceeded $2$ (range: $0.2$–$1.5$). The person separation index or reliability coefficient was $0.59$. None of the 30 items showed DIF with significance ($P > 0.05$) by sex or age.

There was a significant correlation between the person parameter and the arithmetic sum of visual disability score ($R = 0.42$, $P < 0.001$). The item parameters distributed between $−0.88$ and $2.06$, indicating that the items cover a wide range of difficulty; however, the items did not cover patients with mild glaucomatous deterioration.

Figure 2a illustrates the relationship between the mTD of a patient’s IVF and the Rasch analysis–derived person parameter as well as the arithmetic visual disability score (Fig. 2b). There was a significant negative relationship between the Rasch–derived person parameter and the mTD of the IVF ($r = −0.78$, $P$...
FIGURE 1. Person–item map. In the person–item map, items listed closer to the top are more difficult to perform; moving down the scale, the items become easier. The logit scale for item and person parameters is listed across the horizontal axis at the bottom of the map. The distribution of person parameters is shown as the column graph at the top of the map, and the item parameters are shown as the horizontal segment of lines with the two thresholds of response category (indicated as ○) and item difficulty (indicated as ●) for each item. 1 represents the threshold between score 0 and 1, and 2 represents the threshold between the scores of 1 and 2.

< 0.001), which was significantly higher (Meng-Rosenthal-Rubin method, \( P = 0.002 \)) than the correlation between the arithmetic sum of visual disability score and the mTD of the IVF (\( R = -0.61 \), \( P < 0.001 \)). Similarly, the Rasch-derived person parameter showed a significant negative relationship with better- or worse-eye VAs (\( R = -0.58 \) and \(-0.67 \), respectively, \( P < 0.001 \)), which was significantly stronger (Meng-Rosenthal-Rubin method, \( P < 0.001 \)) than the correlation between arithmetic sum of visual disability score and better- or worse-eye VAs (\( R = -0.34 \) and \(-0.40 \), respectively, \( P < 0.001 \)).

**DISCUSSION**

In this study, the validity of the Sumi VDQ was investigated using Rasch analysis. Constructive psychometric properties were observed without re-engineering the original questionnaire. In addition, significantly higher correlations were observed between the Rasch-derived person parameter and the mTD of patients’ IVF and patients’ VAs, compared to the arithmetic sum of visual disability score and these clinical measurements.

Previously, the Sumi VDQ has been reported to possess acceptable reliability (Cronbach’s \( \alpha \), 0.85–0.88), our results here are in strong agreement, demonstrating an even larger Cronbach’s \( \alpha \) statistic equal to 0.96. Rasch analysis revealed an infit value outside the acceptable range of 0.7 to 1.3 for just three items (item 1: 0.55, item 10: 1.44, item 22: 0.68); however, these values were in the range of 0.5 to 1.5, which has been reported as still constructive. Furthermore, none of the PCA components had eigenvalues whose lower limit of the 95% CI exceeded 2, indicating that the Sumi VDQ possesses unidimensionality.

There was a clear linear relationship between the Rasch-derived person parameter and the mTD of the IVF in patients with early-stage glaucoma (Fig. 2a); these same patients tended to have a wide range of arithmetic visual disability scores (Fig. 2b). Furthermore, the scatter plot associated with the Rasch-derived person parameter is more densely distributed around the regression line in the middle to advanced stage than in the scatter plot with the arithmetic visual disability score. Thus it may be useful to use the person parameter, instead of the arithmetic sum of visual disability, in the clinical setting where clinicians see a wide range of patients with early- to advanced-stage glaucoma.

The Sumi VDQ has a three-category response scale. It has been reported that people tend to use only four or five categories to describe their quality of life, but many other questionnaires use just three categories. The number of categories is important because a low frequency can be problematic, as it may not provide stable threshold values. The person–item map (Fig. 1) suggests that the current Sumi VDQ is not sufficiently stable in an early-stage glaucoma patient, because none of the items reached an item parameter value of \(-2 \), whereas the person parameter distributed to \(-4 \). Future studies should investigate whether increasing the number of category responses to four or five can resolve this problem. In addition, the mean of the person parameter was much smaller than the mean of the item parameter (2.17 logits), which represents significant mistargeting. This suggests that a large number of participants in the current study had moderate- to late-stage glaucoma. Thus, it would be useful to further assess the Sumi VDQ in a larger number of patients with early-stage glaucoma.

In the current study, the mean item parameter, which is usually set at zero, was found to equal 0.52 in the Sumi VDQ. This indicates that the set of items is appropriate; however, the person parameter ranged from 4.50 to 3.02, indicating that a high number of patients with relatively preserved visual disability were included in the current study. It should be
Figure 2. (a) The relationship between the Rasch analysis–derived person parameter and the mean of the total deviation values (mTD) of the integrated visual field (IVF). The regression line was expressed as $Y = -3.6 \times X - 17.1$, $R = -0.78$ ($P < 0.001$). (b) The relationship between the arithmetic sum of visual disability score and mTD. The regression line was expressed as $Y = -0.37 \times X - 6.21$, $R = -0.61$ ($P < 0.001$). $R$, Pearson’s correlation coefficient.

further investigated whether the current results are still applicable in a more advanced glaucoma population.

In Japan, sentences are traditionally read and written vertically, so it is highly likely that this custom has influenced the relationship between the VF and visual disability scores in the tasks of “letters” and “sentences.” The Sumi VDQ is the only visual disability instrument to assess visual disability in patients who read and write vertically. More than 1.4 billion people write and read vertically, most commonly in East Asia, as in Hong Kong, Mongolia, Taiwan, and China. Thus the Sumi VDQ is especially useful in countries in this region.

A limitation of the Sumi VDQ is that it does not include items about emotional well-being, social relationships, and independence, which are recommended to be included in visual disability instruments in addition to the visual disability questions included in the Sumi VDQ. Also, an item on driving is not included, because many patients do not commonly drive in Tokyo. A further study should be carried out to incorporate these items in an attempt to improve the questionnaire. Finally, mTD, not pattern standard deviation (PSD), was used to evaluate the Sumi instrument’s consistency with established measurements of VF in the current study. Early glaucomatous VF change may be better reflected using PSD as opposed to mTD values, but PSD decreases when the glaucoma progresses beyond a moderate stage; hence, we chose mTD in the current study. Nonetheless, it may be advantageous to use PSD when validating questionnaires in glaucoma patients with early- to moderate-stage disease.

In conclusion, the results of the Rasch analysis conducted here suggest that the Sumi Visual Disability Questionnaire has constructive psychometric properties and is a valid instrument to assess visual disability in glaucoma patients with a wide range of visual defects.

Acknowledgments

Supported in part by Japan Science and Technology Agency (JST) CREST (RA, HM) and Grants 60645000 (HH), 25861618 (HM), 50570701 (CM), and 26462679 (RA) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Disclosure: H. Hirasawa, None; H. Murata, None; C. Mayama, None; R. Asaoka, None

References


