



Archived at the Flinders Academic Commons:

<http://dspace.flinders.edu.au/dspace/>

'This is the peer reviewed version of the following article:
Jyoti Khadka, Konrad Pesudovs, Colm McAlinden, Michaela
Vogel, Marcus Kernt, Christoph Hirneiss; Reengineering the
Glaucoma Quality of Life-15 Questionnaire with Rasch
Analysis. Invest. Ophthalmol. Vis. Sci.
2011;52(9):6971-6977. doi: 10.1167/iovs.11-7423.

which has been published in final form at

<http://dx.doi.org/10.1167/iovs.11-7423>

This article is made available with the permission of the
publisher, Association for Research in Vision and
Ophthalmology

Copyright © Association for Research in Vision and
Ophthalmology

Reengineering the Glaucoma Quality of Life-15 Questionnaire with Rasch Analysis

Jyoti Khadka,¹ Konrad Pesudovs,¹ Colm McAlinden,¹ Michaela Vogel,² Marcus Kernt,² and Christoph Hirneiss²

PURPOSE. To investigate, using Rasch analysis, whether the 15-item Glaucoma Quality of Life-15 (GQL-15) forms a valid scale and to optimize its psychometric properties.

METHODS. One hundred eighteen glaucoma patients (mean age, 65.7 years) completed the German-version of the GQL-15. Rasch analysis was performed to assess category function (how respondents differentiated between the response options), measurement precision (discriminative ability), unidimensionality (whether items measure a single construct), targeting (whether items are of appropriate difficulty for the sample), and differential item functioning (whether comparable subgroups respond differently to an individual item). Where any of these attributes were outside acceptable ranges, steps were taken to improve the instrument.

RESULTS. The five-response categories of the GQL-15 were well differentiated by respondents, as demonstrated by ordered and well-spaced category thresholds. The GQL-15 had an excellent measurement precision but demonstrated poor targeting of item difficulty to person ability and multidimensionality, indicating that it was measuring more than one construct. Removal of six misfitting items created a nine-item unidimensional instrument with good measurement precision and no differential item functioning but poor targeting. A new name, the Glaucoma Activity Limitation (GAL-9) questionnaire, is proposed for the short version, which better reflects the construct under measurement.

CONCLUSIONS. The GAL-9 has superior psychometric properties over the GQL-15. Its only limitation is poor targeting of item difficulty to person ability, which is an inevitable attribute of a vision-related activity limitation instrument for glaucoma patients, most of whom have only peripheral visual field defects and little difficulty with daily activities. (*Invest Ophthalmol Vis Sci.* 2011;52:6971–6977) DOI:10.1167/iops.11-7423

Glaucoma is the second leading cause of blindness after cataracts; it affects approximately 68 million people worldwide, 10% of whom are blind.^{1,2} Glaucoma is often asymptomatic and therefore recognized as the silent cause of

blindness. Nevertheless, the diagnosis, which requires lifelong follow-up and frequent ocular antihypertensive medication or surgery, can have a huge impact on a patient's life.³ Hence, a comprehensive assessment of the impact of the disease and its treatment on patients from their perspective has become important for the measurement of glaucoma impact and treatment outcomes. The patient's point of view is measured using various types of questionnaires known as patient-reported outcomes (PROs). Highlighting their importance, the US Food and Drug Administration has also endorsed that PRO measures be included in all clinical trial end points for disease impact and outcome assessment in glaucoma.⁴

A number of glaucoma-specific questionnaires or instruments have been developed in the past two decades.^{5–10} The Glaucoma Quality of Life-15 (GQL-15) questionnaire is concise and easy to administer.^{11,12} Independent reviews have described it as one of the better glaucoma-specific instruments, with good acceptability among clinicians and patients.^{13,14} Initially derived from a 62-item pilot instrument, the 15 item GQL-15 was first described in 2003.¹² These 15 items were selected on the basis of their strong relationship with visual field loss in glaucoma patients.¹¹ Several studies have used the GQL-15.^{15–17} The name of the instrument suggests that the trait under measurement is vision-related quality of life; however, all the items refer to activity limitation (near vision, peripheral vision, mobility, and dark adaptation).¹¹

Although there is no universally accepted definition of vision-related quality of life, there is growing consensus that it should include multidimensional assessment of the impact of vision on everyday activities, emotional well-being, social relationships, and independence.¹⁸ The World Health Organization's International Classification of Functioning (WHO-ICF) also provides a unifying framework for the health-related consequences of a disease based on three components: impairment, activity limitations, and participation restriction.¹⁹ When the WHO-ICF framework is conceptualized in vision, impairment refers to the diseases and disorders of the eye (e.g., glaucoma). Activity limitations are the difficulties in executing vision-related tasks as a result of impairment, such as inability to cross the road because of glaucomatous visual field loss. Participation restrictions are barriers to involvement in life situations caused by activity limitation, such as inability to go shopping because of inability to cross the road. When following the WHO-ICF framework, the construct being measured by the GQL-15 is vision-related activity limitations, not quality of life. Indeed, the papers describing its development explain that the purpose is to measure self-reported visual disability despite the name containing the term quality of life.¹¹

Similar to other glaucoma-specific instruments, the GQL-15 was developed and validated using the traditional method (Classical Test Theory [CTT]).¹¹ CTT provides a limited assessment of the psychometric properties of an instrument and produces scoring by the sum of raw ordinal values assigned to each item, which is not a true interval-level measurement.^{20,21}

From the ¹National Health and Medical Research Council (NH&MRC) Centre for Clinical Eye Research, Discipline of Optometry and Vision Science, Flinders Medical Centre and Flinders University of South Australia, South Australia, Australia; and ²Ludwig-Maximilians-University, Department of Ophthalmology, Klinikum der Universität München Campus, Munich, Germany.

Submitted for publication February 20, 2011; revised June 5, 2011; accepted July 21, 2011.

Disclosure: **J. Khadka**, None; **K. Pesudovs**, None; **C. McAlinden**, None; **M. Vogel**, None; **M. Kernt**, None; **C. Hirneiss**, None

Corresponding author: Konrad Pesudovs, NH&MRC Centre for Clinical Eye Research, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, South Australia, 5042, Australia; konrad.pesudovs@flinders.edu.au.

This limits the interpretability of the instrument as a measure.²² The problem can be resolved with the use of Rasch analysis. Rasch analysis estimates the raw questionnaire data to interval-level data.²³ It also provides greater insight into the psychometric properties of an instrument, including the assessment of response categories, measurement of precision, item fit to the construct, item targeting, and unidimensionality. Interval-level data not only provide a valid measurement, they enable the use of robust parametric statistics.²² These benefits of improved psychometric assessment and interval-level scoring have led to the use of Rasch analysis in the development of new questionnaires²⁴⁻²⁶ and the reengineering of existing questionnaires.²⁷⁻³⁰ To the authors' knowledge, the Glaucoma Symptom Scale (GSS), is the only glaucoma-specific instrument that was Rasch analyzed, but it did not demonstrate to have satisfactory psychometric properties.³¹

The primary aim of the present study was to explore the psychometric properties of the GQL-15 using Rasch analysis and to assess whether it forms a valid scale. If the GQL-15 was found to form a valid scale but to have suboptimal psychometric properties, the secondary aim was to optimize its psychometric properties.

METHODS

GQL-15

All the items in the GQL-15 (Table 1) are scored on a five-category difficulty scale, as follows: 1 = no difficulty, 2 = a little bit of difficulty, 3 = some difficulty, 4 = quite a lot of difficulty, 5 = severe difficulty. An additional category (Do not perform for nonvisual reasons) was scored as missing data for the final analysis.

The original GQL-15 was developed in English. To use it with German-speaking patients, it was translated to German (Supplementary File S1, <http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7423/-DCSupplemental>) using a forward and backward translation protocol, with attention paid to maintain the actual meaning of each question.

Patients

The study population consisted of patients who came for regular follow-up in the glaucoma unit of the Department of Ophthalmology at Ludwig-Maximilians-University in Munich, Germany. The German version of the GQL-15 was self-administered by the patients in the clinic. Patients were 18 years of age or older, spoke German, and had no cognitive impairment. Patients with all stages of glaucomatous optic

nerve damage and undergoing all types of treatment, including surgery, were included. Visual acuity in the better eye had to be at least 0.6 logMAR (6/24). Clinical assessment, including the measurement of visual acuity and visual field of each eye, was performed during the follow-up visit. Visual field was assessed using the Humphrey Field Analyzer, SITA 30-2 standard, generating the mean deviation (MD) and the pattern SD in decibels (dB). Severity of glaucoma in the study population was classified on the basis of MD in the visual field as early (-0.01 to -6.00 dB), moderate (-6.01 to -12.00 dB) and advanced/severe (≥ 12.01 dB) glaucoma.³² Ethical approval of the study was obtained from the Institutional Review Board of the University Eye Hospital Munich in Germany, and each patient who agreed to participate signed a consent form. The study adhered to the tenets of the Declaration of Helsinki.

Rasch Analysis

The data were analyzed in two phases: assessment of the psychometric properties of the original GQL-15 and reengineering of the GQL-15 to optimize its psychometric properties.

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. A logit (log-odds ratio) is an interval scale that represents the probability of a person endorsing a particular response category in an item over $(1 - \text{the same probability [i.e., } \ln(p/1 - p)])$. For this analysis, the person with higher ability and items of greater difficulty were located on the negative side of the logit scale and vice versa. Rasch analysis was used for the following assessments: response scale analysis, measurement precision, unidimensionality, targeting, and differential item functioning.

Response Scale Analysis. Whether the response categories were used by respondents in the order intended was evaluated by observing whether the category calibration increased in an orderly fashion in the category probability curves (a graphical display of the likelihood of each category being selected over the range of the scale). 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 GQL-15 are scored on a five-category response scale of increasing difficulty and, therefore, have four thresholds. Disordering of the threshold can occur when a category is underused, its definition is unclear, or the number of categories exceeds the number of levels respondents can distinguish.³³ Disordered thresholds are a source of item misfitting. Therefore, if disordered categories are found, they should be repaired by combining adjacent categories until the thresholds are ordered.³⁴ In addition, categories should be evenly spaced and advance step calibrations by at least 1.4 logits.³⁵

Measurement Precision. In Rasch analysis, measurement precision of an instrument is denoted by person separation. This is a measure of discriminative ability of an instrument. A person separation of 2.00 indicates that at least three strata or groups of person ability can be discriminated. The higher the person separation, the more groups the instrument can distinguish. A person separation of 2.00 is the minimum accepted level of discrimination for an instrument to produce a valid measure.³⁵

Unidimensionality. A fundamental element of measurement is unidimensionality; a score produced by a measure should represent a single concept. Unidimensionality is assessed by examining the fit statistics and principal component analysis (PCA) of the residuals.³⁶ There are two types of fit statistics, infit and outfit. Both statistics indicate how well items fit the underlying construct. Infit is more sensitive to unexpected responses to items by participants whose ability is near (inliers) item difficulty level, whereas outfit is more sensitive to unexpected responses to items by participants whose ability is far (outliers) from item difficulty level. Therefore, infit is considered more informative.³⁴ Fit statistics are measured as mean square standardized residuals (MNSQ). The standard cutoff range for MNSQ is from 0.7 to 1.3,³⁷ but values between 0.5 and 1.5 could be considered productive for measure-

TABLE 1. The GQL-15 Questionnaire

"Does your vision give you any difficulty, even with glasses, with the following activities?" (1 = no difficulty, 2 = a little bit of difficulty, 3 = some difficulty, 4 = quite a lot of difficulty, 5 = severe difficulty)

1. Reading newspapers
2. Walking after dark
3. Seeing at night
4. Walking on uneven ground
5. Adjusting to bright lights
6. Adjusting to dim lights
7. Going from light to dark room or vice versa
8. Tripping over objects
9. Seeing objects coming from the side
10. Crossing the road
11. Walking on steps/stairs
12. Bumping into objects
13. Judging distance of foot to step/curb
14. Finding dropped objects
15. Recognizing faces

TABLE 2. Sociodemographic and Visual Characteristics of the 118 Glaucoma Patients

Variables	Results
Mean age \pm SD, y	65.71 \pm 13.63
Age range, y	23 to 88
Sex, female, <i>n</i> (%)	71 (60.2)
Better eye visual acuity	
LogMAR	
Mean \pm SD	0.11 \pm 0.13
Range	-0.13 to 0.52
Median (interquartile range)	0.10 (0.00 to 0.22)
Snellen	
Mean	6/7.5
Range	6/5 ⁺² to 6/20 ⁻¹
Worse eye visual acuity	
LogMAR	
Mean \pm SD	0.35 \pm 0.40
Range	0.00 to 3.00
Median (interquartile range)	0.22 (0.1 to 0.49)
Snellen	
Mean	6/12 ⁻²
Range	6/6 to 6/6000
Better eye visual field	
Mean deviation, mean \pm SD, db	-4.68 \pm 5.75
Pattern standard deviation, mean \pm SD, db	4.37 \pm 3.21
Worse eye visual field	
Mean deviation, mean \pm SD, db	-7.87 \pm 6.37
Pattern standard deviation, mean \pm SD, db	6.13 \pm 3.68

ment.³⁸ Ideally, item fit <0.7 indicates redundancy, and values higher than 1.3 indicate unacceptable levels of noise in the responses and are considered misfitting. Misfitting items should be removed to reduce noise and to optimize the psychometric properties of an instrument. The removal of misfitting items should occur one at a time until all the remaining items demonstrate good fit.

Unidimensionality was further assessed by PCA of the residuals. If the principal component of the residuals explains 60% or greater variance of the model, then there is high likelihood that an instrument is measuring a single underlying trait or is unidimensional. The contrasts in the residuals report unexplained variance by the principal component. This study used the criterion that the contrast with an eigenvalue >2.00 U (i.e., strength of at least two items) is suggestive of a second construct being measured, thus indicating a multidimensional instrument. Similarly, items loading on first contrast by a minimum of 0.4 are identified as contrasting items and tap different constructs.

Targeting. Rasch analysis generates a person-item map that provides a visual observation of the relative position of item difficulty to person ability. By default, the item mean is placed at 0 logit. For a perfectly targeted instrument, both item and person means lie on the sample point on the map (i.e., mean difference = 0 logits). However, a difference of person and item means of up to 1 logit is acceptable. A difference between means of >1 logit indicates notable mistargeting. Poor targeting occurs because of items clustering at a certain point along the map, large gaps between items, and the higher or lower ability of the study population than the required level of ability to endorse the items.

Differential Item Functioning. Rasch analysis permits testing of differential item functioning (DIF). DIF occurs when different subgroups of people with comparable levels of ability respond differently to an item. DIF magnitude of <0.5 logit is considered small or absent, 0.50 to 1.0 logit is minimal, and >1.0 logit is notable.³⁹ For DIF testing, the study population was stratified by sex and age (younger than 65 years and 65 years or older).

Validity. Validity refers to the assessment of whether an instrument measures what it purports to measure. Validity of the instrument was assessed by calculating correlation coefficient between the questionnaire score and traditional clinical measures (that is, visual acuity and visual field). Such a validity is called criterion or external validity,

which describes whether the measures of an instrument are consistent with the established measurements.³⁵

Statistical Analysis

Rasch analysis was performed on the GQL-15 questionnaire data according to the Andrich Rating scale model using joint likelihood estimation with the Winsteps (version 3.67.0; SAS Institute, Chicago IL) software.⁴⁰ The Predictive Analytics Software (PASW, version 18.0; SAS Institute) was used for all general descriptive statistics. Associations between normally distributed data were described by Pearson correlation coefficient, whereas Spearman rank correlation was used if one datum or both data were not distributed normally. $P < 0.05$ (two-tailed) was considered statistically significant.

RESULTS

One hundred eighteen glaucoma patients completed the German version of the GQL-15 by self-administration. Table 2 shows the sociodemographic and visual characteristics of the participants. On the basis of MD in better eye, 72.4%, 16.3%, and 11.3% of the study population were classified as having early, moderate, and advanced forms of glaucoma. When considering worse eyes, 50%, 26.1%, and 23.9% had early, moderate, and advanced glaucoma.

Assessment of the Psychometric Properties of the Original GQL-15

Response Scale Analysis. The GQL-15 instrument demonstrated a well-functioning response scale. There were ordered thresholds between all response categories, indicating that each category had distinct meaning. The spacing of each category on the scale was even (>1.4), illustrating that each category had equal probability to be endorsed by the participants. The percentages of category utilization were 44% (1 = no difficulty), 28% (2 = a little bit of difficulty), 17% (3 = some difficulty), 7% (4 = quite a lot of difficulty), and 4% (5 = severe difficulty). Figure 1 shows the ordered response options with a distinct peak for each response category.

Measurement Precision and Targeting. The person separation value for the GQL-15 was excellent (3.14), which indicates that the instrument can distinguish at least four groups of

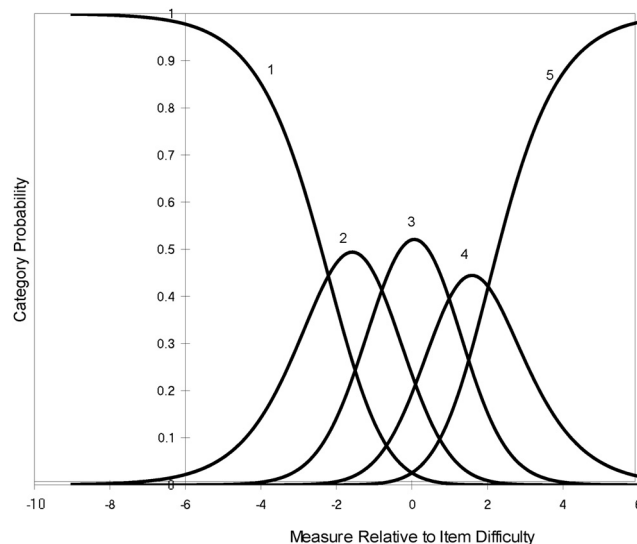


FIGURE 1. Category probability curves showing five well-functioning categories: 1 = no difficulty; 2 = a little bit of difficulty; 3 = some difficulty; 4 = quite a lot of difficulty; and 5 = severe difficulty.

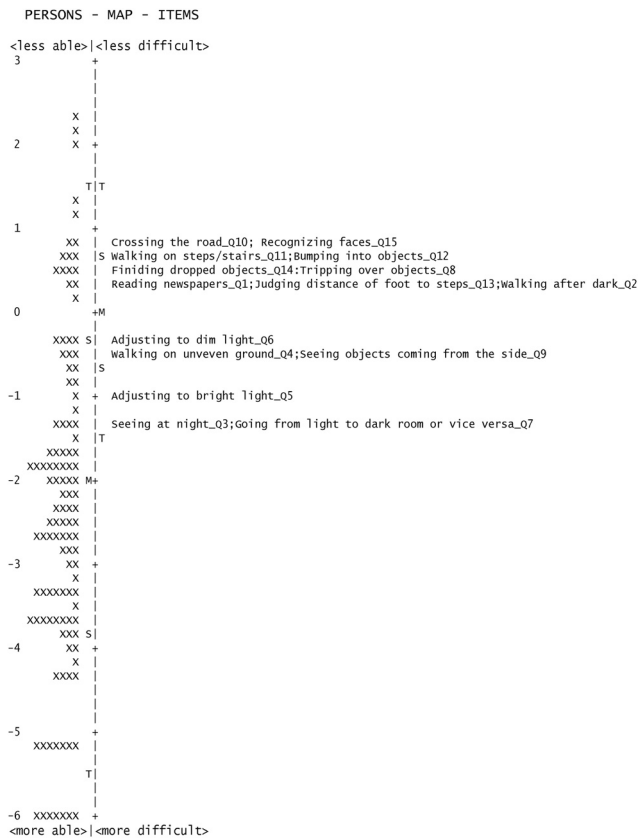


FIGURE 2. Person-item map for the GQL-15. The participants are represented by x on the left, and items are located on the right of the dashed line. The more difficult items and more visually able participants are located at the bottom of the map. Each x represents one participant. M, mean; S, 1 SD from the mean; T, 2 SD from the mean.

person ability. In Figure 2, the spread of each item calibration is visualized compared with the range of person ability estimates. The items were not targeted well to persons' ability in the sample, which was demonstrated by a large gap between person and item mean (-2.07).

Item Fit and Unidimensionality. All 15 items fit the Rasch model within liberal infit (0.66-1.41) and outfit (0.59-1.39) ranges and, hence, could be considered productive for

measurement. However, those items outside the range of 0.7 to 1.3 were a potential source of noise and could be considered for removal to optimize the psychometric properties of the instrument. Two items (1 and 15) were underfitting, and three items (8, 12, and 10) were overfitting the Rasch model (Table 3).

The PCA of the residuals showed that the variance explained by the principal component was 65.6%. However, the unexplained variance explained by the first contrast was 2.2 eigenvalue units, which suggests that the instrument was not unidimensional. Three items representing mobility (10, Crossing the road; 12, Bumping into objects; and 11, Walking on steps/stairs) loaded positively by >0.4 onto first contrast. Similarly, two items representing dark adaptation (5, Adjusting to bright lights; and 7, Going from light to dark room or vice versa) loaded negatively by <-0.4 onto the first contrast. No further contrast exceeded 2.0 eigenvalue units.

Differential Item Functioning. None of the 15 items showed DIF by age or sex (Table 3).

Reengineering the GQL to Optimize Its Psychometric Properties. The main problem identified in the Rasch analysis of the GQL-15 is multidimensionality, illustrated by the PCA of the residuals and supported by item fit statistics. The elimination of multidimensionality is essential to the validity of this instrument; therefore, item removal was undertaken. The items in the first PCA contrasts and the misfitting items were removed one by one until all items had fit statistics ranging from 0.7 to 1.3 and no contrast had >2.0 eigenvalue units. The five items that initially misfit (items 1, 8, 10, 12 and 15) were removed one at a time. Item 5 misfitted (infit, 1.33; outfit, 1.26) the Rasch model after removal of the above five items and was also removed. This process provided an acceptable fit statistics of the remaining items (Table 4) and established unidimensionality in the scale. The PCA showed that the principal component explained 68.6% of the modeled variance, and no contrast exceeded 2.0 eigenvalue units. This strongly suggests that the nine-item instrument is unidimensional. The shorter nine-item instrument had good person separation (2.66), and targeting was slightly better (-1.90) than that of the longer version (Fig. 3). Item 3 (Seeing at night) was the most difficult item or high-level ability discriminating item. Conversely, item 11 (Walking on steps/curbs) discriminated less able people. The item measure of the scale ranged from 1.05 to -1.22 logit. As in the longer version, the remaining nine items were also free of DIF by age and sex. The removed items were tested for fit to the Rasch model to determine whether they could measure an additional construct, but the mea-

TABLE 3. Rasch Fit Statistics and Item Measure for the GQL-15

Item No.	Items	Infit (MNSQ)	Outfit (MNSQ)	Item Measure (logit)
1	Reading newspapers	1.41	1.39	0.28
2	Walking after dark	1.05	0.80	0.33
3	Seeing at night	1.13	0.08	-1.35
4	Walking on uneven ground	0.91	0.83	-0.43
5	Adjusting to bright lights	1.25	1.25	-0.97
6	Adjusting to dim lights	1.10	1.08	-0.34
7	Going from light to dark room or vice versa	1.03	0.99	-1.31
8	Tripping over objects	0.66	0.59	0.49
9	Seeing objects coming from the side	1.02	1.07	-0.45
10	Crossing the road	0.70	0.61	0.78
11	Walking on steps/stairs	0.75	0.91	0.75
12	Bumping into objects	0.68	0.55	0.63
13	Judging distance of foot to step/curb	0.83	0.98	0.28
14	Finding dropped objects	0.88	0.80	0.49
15	Recognizing faces	1.36	1.22	0.82

Items with fit statistics outside the acceptable range appear in bold.

TABLE 4. Rasch Fit Statistics and Item Measure for the 9-Item and 5-Response Category Instrument

“Does your vision give you any difficulty, even with glasses, with the following activities?”
(1 = no difficulty, 2 = a little bit of difficulty, 3 = some difficulty, 4 = quite a lot of difficulty, 5 = severe difficulty)

GQL-15 Item Number	New Item Number	Items	Infit (MNSQ)	Outfit (MNSQ)	Item Measure
2	1	Walking after dark	1.07	0.81	0.61
3	2	Seeing at night	1.01	0.96	-1.22
4	3	Walking on uneven ground	0.83	0.78	-0.22
6	4	Adjusting to dim lights	1.10	1.16	-0.13
7	5	Going from light to dark room and vice versa	1.03	1.00	0.61
9	6	Seeing objects coming from the side	1.09	1.13	-0.24
11	7	Walking on steps/stairs	0.97	0.86	1.05
13	8	Judging distance of foot to step/curb	0.89	0.97	0.55
14	9	Finding dropped objects	1.01	0.93	0.78

surement precision was inadequate (person separation, 1.87).

Validity Assessment of the Nine-Item Questionnaire.

Given that MD and the nine-item questionnaire score (person estimate in logit) were normally distributed, Pearson's correlation coefficient was used to calculate correlation between them. There was a significant negative correlation of the instrument score with the worse eye MD ($r = -0.453$; $P < 0.001$), and a weak correlation was observed with the better eye MD ($r = -0.296$; $P = 0.04$). Because visual acuity was not distributed normally, Spearman's rank test was used to calculate correlation with the Glaucoma Activity Limitation-9 (GAL-9) score. A medium association was also observed between questionnaire score and visual acuity, and it was slightly higher with the eye worse eye (0.418; $P < 0.001$) than with the better eye (0.323; $P = 0.001$).

DISCUSSION

This study shows that the GQL-15 functions within the Rasch model in terms of measurement precision, response category functioning, and DIF. However, it was not a unidimensional scale. This is a fundamental problem because it becomes unclear what the instrument measures. To draw a clinical analogy, imagine a device that measures both intraocular pressure (IOP) and central corneal thickness but produces only one score. What would 600 mean, a thick cornea and a low IOP or a thin cornea and a high IOP? For glaucoma management, it is imperative that the two constructs under measurement be segregated into identifiable components. The same is true in questionnaires. Our analysis suggests that the GQL-15 largely measures one construct (vision-related activity limitation) but is contaminated by a mobility construct that appears to be different. Creation of a nine-item version enabled unidimensional measurement with excellent psychometric attributes. Hence, the shorter version is a better measure than the GQL-15 of vision-related activity limitation in patients with glaucoma.

To establish unidimensionality in the GQL-15, six items (1, 5, 8, 10, 12, and 15) were considered for removal on the basis of the PCA of the residuals and fit statistics. Among the three items related to mobility, the two items (10, crossing the road; 12, bumping into objects) loaded >0.4 in the first PCA contrast, and one item (8, tripping over objects) grossly misfitted the Rasch model. These activities may be important in glaucoma patients,⁴¹ but Rasch analysis identified these items tap a different construct so are as a source of noise and multidimensionality in this activity limitations scale. Similarly, two near vision items (1, reading the newspaper; 15, recognizing faces)

misfitted the model, possibly for a number of reasons. For example they may be related to near vision correction rather than to glaucoma; hence, these items do not behave predictably across the whole population. A glare disability item (5, adjusting to bright light) also misfitted the Rasch model. This was perhaps influenced by the presence of cataract or innate photophobia in a subset of patients; thus, the item did not behave as expected and again was removed from the scale.

Interestingly, though three mobility items misfitted and were removed, the shorter version still contains four mobility items (2, 4, 11, and 13). Similarly, three items that are related to light and dark adaptation (3, seeing at night; 6, adjusting to dim lights; 7, going from a light room to a dark room or vice versa) are also retained in the short version. Although it may seem inconsistent to remove some items and retain others within these two conceptual areas, the key issue is that the items retained behave predictably within the whole item set across the entire glaucoma population whereas the removed items did not.

The GQL has a five-category response scale that performs well. The categories are used in the order intended, and each occupies a wide portion of the measurement scale. The only concern is that the two higher-end categories had a low frequency of utilization (5, severe difficulty = 4%; 4, a lot of difficulty = 7%). Categories with low frequency can be problematic because they do not provide stable threshold values, and it may be recommended that they be collapsed into adjacent categories to eliminate noise that may arise from unstable calibrations.³⁴ We could have collapsed these two categories to improve utilization frequency. However, this would not have improved the psychometric properties of the instrument. As is typical of glaucoma patients, the majority of the study population had low visual disability. It is simply unlikely that many patients with glaucoma would endorse the categories of quite a lot of difficulty or severe difficulty. However, removing these categories may lead to loss of valuable psychometric information when the instrument is used on patients with higher disability.⁴² Hence, five response categories were retained.

A significant association between instrument score and visual field loss was found in this study. The impact of visual field loss was stronger in the worse eye than in the better eye on the (GAL-9) score. Other studies have also reported that visual field loss in the worse eye has a great influence in self-reported visual disability in patients with glaucoma.^{3,17,43} Similarly, visual acuity in the worse eye also demonstrated better association with the GAL-9 score than visual acuity in the better eye. However, the association was not distinctly different as in the

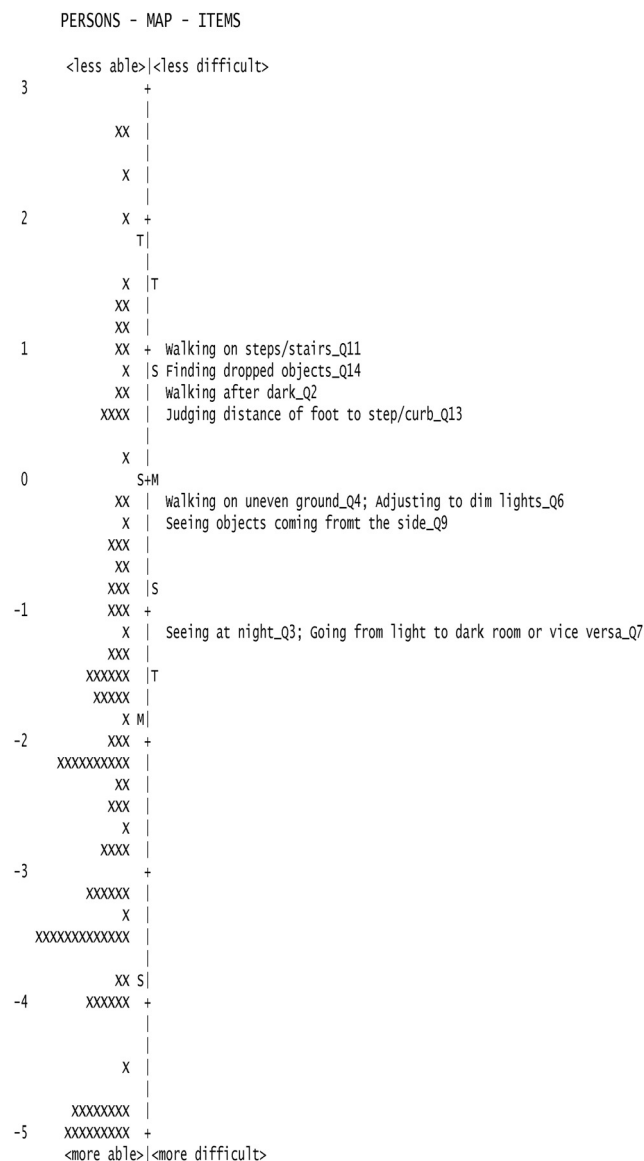


FIGURE 3. Person-item map for the GAL-9. The participants are represented by *x* on the left, and items are located on the right of the dashed line. The more difficult items and more visually able participants are located at the bottom of the map. Each *x* represents one participant. M, mean; S, 1 SD from the mean; T, 2 SD from the mean.

worse and better eye visual field loss. Significant correlations between the GAL and visual parameters have demonstrated its validity. It would also have been interesting to investigate the correlation between the binocular visual fields on the questionnaire score. However, the binocular visual field test was not recorded in our study population.

The original name of the questionnaire might be misleading because the instrument measures only vision-related activity limitations. Hence, we have proposed a new name for the short-version instrument, the GAL-9 questionnaire. The GAL-9 also demonstrates poor targeting (difference between item mean and person mean ≥ 1.00 logit).³⁰ Poor targeting is an inevitable attribute because most people with glaucoma have little difficulty in performing everyday tasks^{44,45} until they have poor visual acuity and advanced visual field loss.^{12,46} This is consistent with the findings of this study and the visual acuity and visual field status of our population. Poor targeting is also a common problem in other vision-specific questionnaires

when used on more able patients, such as in second eye cataract surgery patients who become more able after first eye cataract surgery.^{39,47,48}

Alternatively, more sensitive items addressing higher visual ability can be added to optimize the targeting of the instrument, but this strategy requires revalidation with each new addition. Such a process is lengthy, time consuming, and does not guarantee avoidance of poor targeting. It might be a good idea to develop a new comprehensive instrument that addresses holistic issues such as treatment effects and the psychosocial impact of glaucoma or to develop a superior strategy in the form of item banking. An item bank consists of a larger number of Rasch-calibrated items in a pool, which are presented by computer-adaptive testing (CAT) to patients on the basis of their response to previous items. This tailoring of item presentation ensures targeting of item difficulty to person ability. Hence, an item bank with the use of CAT provides a rapid, precise, and accurate measurement of the impact of disease on patients.⁴⁹

In conclusion, the revised GAL-9 has superior psychometric properties, including unidimensionality and good measurement precision with the added advantage of low respondent burden. Indeed, the GAL-9 is the only PRO designed for use in glaucoma patients to have been demonstrated to have satisfactory psychometric properties (the Glaucoma Symptom Scale does not).³¹ Therefore, we encourage the use of the GAL-9 because of its psychometric properties and interval scaling, at least until a superior questionnaire is available. To simplify implementation, a spreadsheet (Excel; Microsoft, Redmond, WA) enabling estimation of person ability in logits from category responses is available for download (Supplementary File S2, <http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7423/-DCSupplemental>).

Acknowledgments

The authors thank Pia Thomczyk for participant management and data collection.

References

1. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol*. 2006;90:262–267.
2. Weinreb RN, Kaufman PL. The Glaucoma research community and FDA look to the future: a report from the NEI/FDA CDER Glaucoma Clinical Trial Design and Endpoints Symposium. *Invest Ophthalmol Vis Sci*. 2009;50:1497–1505.
3. Spaeth G, Walt J, Keener J. Evaluation of quality of life for patients with glaucoma. *Am J Ophthalmol*. 2006;141:3–14.
4. Varma R, Richman EA, Ferris FL, Bressler NM. Use of patient-reported outcomes in medical product development: a report from the 2009 NEI/FDA Clinical Trial Endpoints Symposium. *Invest Ophthalmol Vis Sci*. 2010;51:6095–6103.
5. Atkinson M, Stewart W, Fain J, et al. A new measure of patient satisfaction with ocular hypotensive medications: The Treatment Satisfaction Survey for Intraocular Pressure (TSS-IOP). *Health Qual Life Outcomes*. 2003;1:67.
6. Barber BL, Strahlman ER, Laibovitz R, Guess HA, Reines SA. Validation of a questionnaire for comparing the tolerability of ophthalmic medications. *Ophthalmology*. 1997;104:334–342.
7. Béchetolle A, Arnould B, Bron A, et al. Measurement of health-related quality of life with glaucoma: validation of the Glau-QoL[®] 36-item questionnaire. *Acta Ophthalmol*. 2008;86:71–80.
8. Janz NK, Wren PA, Lichten PR, et al. Quality of life in newly diagnosed glaucoma patients: The Collaborative Initial Glaucoma Treatment Study. *Ophthalmology*. 2001;108:887–897.
9. Lee BL, Gutierrez PRMA, Gordon M, et al. The Glaucoma Symptom Scale: a brief index of glaucoma-specific symptoms. *Arch Ophthalmol*. 1998;116:861–866.
10. Nordmann J-P, Denis P, Vigneux M, Trudeau E, Guillemin I, Berdeaux G. Development of the conceptual framework for the

- Eye-Drop Satisfaction Questionnaire (EDSQ(c)) in glaucoma using a qualitative study. *BMC Health Serv Res.* 2007;7:124.
11. Nelson P, Aspinall P, Pappasoulitis O, Worton B, O'Brien C. Quality of life in glaucoma and its relationship with visual function. *J Glaucoma.* 2003;12:139-150.
 12. Nelson P, Aspinall PA, O'Brien C. Patients' perception of visual impairment in glaucoma: a pilot study. *Br J Ophthalmol.* 1999;83:546-552.
 13. Hirneiss C, Kampik A, Neubauer AS. Value-based medicine for glaucoma. *Ophthalmology.* 2010;107:223-227.
 14. Walt JG, Rendas-Baum R, Kosinski M, Vaishali PP. Psychometric evaluation of the Glaucoma Symptom Identifier. *J Glaucoma.* 2011;20:148-159.
 15. Goldberg I, Clement CI, Chiang TH, et al. Assessing quality of life in patients with glaucoma using the Glaucoma Quality of Life-15 (GQL-15) questionnaire. *J Glaucoma.* 2009;18:6-12.
 16. Skalicky S, Goldberg I. Depression and quality of life in patients with glaucoma: a cross-sectional analysis using the Geriatric Depression Scale-15, assessment of function related to vision, and the Glaucoma Quality of Life-15. *J Glaucoma.* 2008;17:546-551.
 17. van Gestel A, Webers CA, Beckers HJ, et al. The relationship between visual field loss in glaucoma and health-related quality-of-life. *Eye.* 2010;24:1759-1769.
 18. Elliott DB, Pesudovs K, Mallinson T. Vision-related quality of life. *Optom Vis Sci.* 2007;84:656-658.
 19. World Health Organization. International Classification of Functioning, Disability and Health (ICF). Geneva: WHO Regional Office for Europe; 2001. http://www.disabilitaincife.it/documenti/ICF_18.pdf. Accessed August 7, 2011.
 20. Mallinson T. Why measurement matters for measuring patient vision outcomes. *Optom Vis Sci.* 2007;84:675-682.
 21. Pesudovs K. Patient-centred measurement in ophthalmology—a paradigm shift. *BMC Ophthalmol.* 2006;6:25.
 22. Svensson E. Guidelines to statistical evaluation of data from rating scales and questionnaires *J Rehab Med.* 2001;33:47-48.
 23. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989;70:857-860.
 24. 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.
 25. McAlinden C, Pesudovs K, Moore JE. The development of an instrument to measure quality of vision: the Quality of Vision (QoV) questionnaire. *Invest Ophthalmol Vis Sci.* 2010;51:5537-5545.
 26. 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.
 27. Gothwal VK, Wright TA, Elliott DB, Pesudovs K. The refractive status and vision profile: Rasch analysis of subscale validity. *J Refract Surg.* 2010;26:912-915.
 28. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Visual Activities Questionnaire: assessment of subscale validity for cataract surgery outcomes. *J Cataract Refract Surg.* 2009;35:1961-1969.
 29. Marella M, Pesudovs K, Keeffe JE, O'Connor PM, Rees G, Lamoureux EL. The psychometric validity of the NEI VFQ-25 for use in a low-vision population. *Invest Ophthalmol Vis Sci.* 2010;51:2878-2884.
 30. Pesudovs K, Gothwal VK, Wright T, Lamoureux EL. Remediating serious flaws in the National Eye Institute Visual Function Questionnaire. *J Cataract Refract Surg.* 2010;36:718-732.
 31. Lamoureux EL, Ferraro JG, Pallant JF, Pesudovs K, Rees G, Keeffe JE. Are standard instruments valid for the assessment of quality of life and symptoms in glaucoma? *Optom Vis Sci.* 2007;84:789-796.
 32. Kymes SM, Kass MA, Anderson DR, Miller JP, Gordon MO. Management of ocular hypertension: a cost-effectiveness approach from the Ocular Hypertension Treatment Study. *Am J Ophthalmol.* 2006;141:997-997.
 33. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas.* 2002;3:85-106.
 34. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Science.* 2nd ed. London: Lawrance Erlbaum Associates; 2007.
 35. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci.* 2007;84:663-674.
 36. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas.* 2002;3:205-231.
 37. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Meas Trans.* 1994;8:370.
 38. Linacre JM. Rasch power analysis: size vs. significance: standardized chi-square fit statistic. *Rasch Meas Trans.* 2003;17:918.
 39. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Rasch analysis of visual function and quality of life questionnaires. *Optom Vis Sci.* 2009;86:1160-1168.
 40. Linacre JM. WINSTEPS Rasch measurement, version 7.70.0. Chicago: Winsteps; 2010.
 41. Iester M, Zingirian M. Quality of life in patients with early, moderate and advanced glaucoma. *Eye.* 2002;16:44-49.
 42. Rogelberg SG, Church AH, Waclawski J, Stanton JM. *Organizational Survey Research: Handbook of Research Methods in Industrial and Organizational Psychology.* Oxford: Blackwell Publishing; 2008:140-160.
 43. McKean-Cowdin R, Wang Y, Wu J, Azen SP, Varma R. Impact of visual field loss on health-related quality of life in glaucoma: the Los Angeles Latino Eye Study. *Ophthalmology.* 2008;115:941-948.
 44. Lamoureux EL, Hassell JB, Keeffe JE. The determinants of participation in activities of daily living in people with impaired vision. *Am J Ophthalmol.* 2004;137:265-270.
 45. Ramulu P. Glaucoma and disability: which tasks are affected, and at what stage of disease? *Curr Opin Ophthalmol.* 2009;20:92-98.
 46. Weih LM, Hassell JB, Keeffe J. Assessment of the impact of vision impairment. *Invest Ophthalmol Vis Sci.* 2002;43:927-935.
 47. Gothwal VK, Wright T, Lamoureux EL, Pesudovs K. Psychometric properties of visual functioning index using Rasch analysis. *Acta Ophthalmol.* 2010;88:797-803.
 48. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Using Rasch analysis to revisit the validity of the Cataract TyPE Spec instrument for measuring cataract surgery outcomes. *J Cataract Refract Surg.* 2009;35:1509-1517.
 49. Pesudovs K. Item banking: a generational change in patient-reported outcome measurement. *Optom Vis Sci.* 2010;87:285-293.