Rasch Analysis of the Daily Living Tasks Dependent on Vision (DLTV)

Frances Denny,[1,2] *Adele H. Marshall,*[2] *Michael R. Stevenson,*[3] *Patricia M. Hart,*[1] *and Usha Chakravarthy*[1]

Purpose. To examine internal consistency, refine the response scale, and obtain a linear scoring system for the visual function instrument, the Daily Living Tasks Dependent on Vision (DLTV).

Methods. Data were available from 186 participants with a clinical diagnosis of AMD who completed the 22-item DLTV (DLTV-22) according to four-point ordinal response scale. An independent group of 386 participants with AMD were administered a reduced version of the DLTV with 11 items (DLTV-11), according to a five-point response scale. Rasch analysis was performed on both datasets and used to generate item statistics for measure order, response odds ratios per item and per person, and infit and outfit mean square statistics. The Rasch output from the DLTV-22 was examined to identify redundant items and for factorial validity and person item measure separation reliabilities.

Results. The average rating for the DLTV-22 changed monotonically with the magnitude of the latent person trait. The expected versus observed average measures were extremely close, with step calibrations evenly separated for the four-point ordinal scale. In the case of the DLTV-11, step calibrations were not as evenly separated, suggesting that the five-point scale should be reduced to either a four- or three-point scale. Five items in the DLTV-22 were removed, and all 17 remaining items had good infit and outfit mean squares. PCA with residuals from Rasch analysis identified two domains containing 7 and 10 items each. The domains had high person separation reliabilities (0.86 and 0.77 for domains 1 and 2, respectively) and item measure reliabilities (0.99 and 0.98 for domains 1 and 2, respectively).

Conclusions. With the improved internal consistency, establishment of the accuracy and precision of the rating scale for the DLTV and the establishment of a valid domain structure we believe that it constitutes a useful instrument for assessing visual function in older adults with age-related macular degeneration. (*Invest Ophthalmol Vis Sci.* 2007;48: 1976–1982) DOI:10.1167/iovs.06-0135

The Daily Living Tasks Dependent on Vision (DLTV) is a quality of life (QOL) questionnaire that was constructed to obtain estimates of self-reported ability to perform vision-related tasks in persons with visual impairment due to age-related macular degeneration (AMD). Composed of 22 items, the instrument contains questions on a variety of daily living tasks relating to vision.[1] The sensitivity of the instrument to visual impairment, its domain structure, reproducibility, and reliability have been described by Hart et al.[1,2]

The DLTV uses a four-point ordinal response scale that reflects the following positions: "Can't see to do" (score, 1), "A lot of difficulty" (score, 2), "A little difficulty" (score, 3), and "No difficulty" (score, 4). We have previously analyzed the data from the DLTV using techniques that assume that the four-point scale is linear and of equal intervals, for example principal component analysis (PCA). More recently, this approach has been questioned,[3] and in this regard, Rasch analysis has become increasingly popular as a measurement for validation or amendment of ordinal scales in QOL questionnaires.[4] Previous work has been undertaken to validate the DLTV as a QOL instrument by using traditional criteria such as correlations with other variables.[2] The principal objectives of the present study were to examine internal consistency and optimize the response scale of the DLTV using Rasch analysis. As Rasch analysis also permits assessment of the content validity of the instrument, this assessment was also undertaken.

Methods

The data were acquired as part of two QOL studies in patients with age-related macular degeneration (AMD). The clinical and QOL outcomes from one of these studies has been reported previously.[1,2,5] The first dataset was collected during the Subfoveal Radiotherapy Study (SFRADS), in which participants with exudative AMD were enrolled in three U.K. study sites from November 1995 to July 1998. A set of QOL instruments, including the DLTV, were administered to the 203 participants enrolled in SFRADS at designated study visits at 0, 6, 12, and 24 months. At the baseline examination, 186 completed DLTV questionnaires were available, constituting the data analyzed in the present report.

A second study was undertaken to examine further aspects of independent living and vision-related QOL in patients with AMD. This study acquired information on the use of health and personal social services and independent living and included 11 items from the DLTV. We designated this subset of items the DLTV-11, and we henceforth refer to the original 22 items as the DLTV-22. The rating scale of the DLTV-11 was modified to an ordinal five-point scale instead of the four-point scale used in the DLTV-22. Three hundred sixty-eight participants who were enrolled in the study between November 2000 and August 2001 were administered the DLTV-11. Complete data were available on 324 of the participants. Any missing data were due to the participant's fatigue or failure to collect the information at the clinic visit.

Both studies had been approved by the Ethics Committee, Queen's University Belfast, and were in accordance with the guidelines in the Declaration of Helsinki on research in human volunteers. All participants gave informed consent after the nature of the study was explained.

Summary statistics for both datasets were generated with statistical analysis software (SPSS ver. 14.0; SPSS, Chicago, IL). Another software package (Winsteps ver. 3.58.1; Winsteps, Chicago, IL) was used to

From the [1]Centre for Vision Sciences and the [2]Centre for Statistical Science and Operational Research, School of Mathematics and Physics, and the [3]Public Health Medicine and Primary Care, Queen's University of Belfast, Belfast Northern Ireland, United Kingdom.

Supported by the Medical Research Council, UK, and The Guide Dogs for the Blind, UK.

Submitted for publication February 7, 2006; revised May 26 and October 16, 2006, and January 8, 2007; accepted March 14, 2007.

Disclosure: **F. Denny**, None; **A.H. Marshall**, None; **M.R. Stevenson**, None; **P.M. Hart**, None; **U. Chakravarthy**, None

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Usha Chakravarthy, Centre for Vision Sciences, Institute of Clinical Science, The Queen's University of Belfast, Belfast BT12 6BA, Northern Ireland, UK; u.chakravarthy@qub.ac.uk.

Investigative Ophthalmology & Visual Science, May 2007, Vol. 48, No. 5

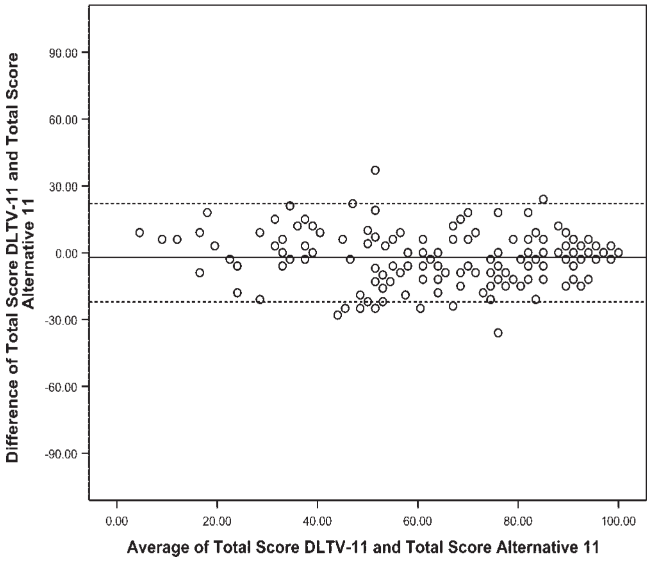

FIGURE 1. Bland-Altman plot of the common 11 items of the DLTV-22 and the DLTV-11 versus the remaining items of the DLTV-22. Averages of the overall score of the 11 items of the DLTV-11 and the remaining 11 items of the DLTV-22 (*x*-axis) were plotted against the differences (*y*-axis). The line $x = -2.1$ is the mean, whereas the *two lines* at ± 22.1 indicate twice the SD.

perform Rasch analysis with the Andrich rating scale model.[6–8] In Rasch analysis, the point at which the probability that a respondent will choose one category over another is known as a threshold. The number of thresholds is one less than the number of categories in the questionnaire and therefore the DLTV-22 instrument has three thresholds and the DLTV-11 has four. At the boundary of each threshold, there is an equal probability of scoring at a lower or higher level. For example, at the boundary of the threshold between "A lot of difficulty" and "Can't see to do," the respondent would be assigned a boundary score of 1. This can be calculated by using a general formula. The odds of choosing a category x over the previous category $x - 1$ is given by

$$\ln(P_{nix}/P_{ni(x-1)}) = B_n - D_i - F_x \text{ the Andrich Rating Scale Model,} \quad (1)$$

where P_{nix} is the probability that person n, on encountering item i would be observed (or would respond) in category x; $P_{ni(x-1)}$ is the probability that the observation (or response) would be in category $x - 1$; B_n is the ability of person n; D_i is the difficulty of item i; and F_x is the threshold for response category x.[6]

Winsteps uses unconditional maximum likelihood estimation routines to estimate latent variable measures for each person (B_n), each item (D_i), and each threshold (F_x). It also provides estimates of fit statistics and separation reliability coefficients for each latent variable.[4]

To examine whether the items that constituted the DLTV-11 and the remaining items of the DLTV-22 were measuring similar constructs, a Bland-Altman plot was generated and examined for agreement. This analysis was accomplished with the SFRADS dataset.

To assess the rating scale, the category frequency counts from the two instruments were examined and probability characteristic curves were generated.

The average raw scores for each individual were plotted against the person measures (B_n) to assess whether the summary score of the DLTV-22 was linear. Person-item maps were generated to examine content validity of the DLTV-22 after separating participants into two groups by level of visual function: group 1 distance visual acuity (DVA) $\geq$ 0.3 logMAR (logarithm of the minimum angle of resolution) in the better-seeing eye ($n = 117$) and group 2 DVA $\leq$ 0.2 logMAR in at least one eye ($n = 82$).

To improve content validity, the Rasch output from the DLTV-22 was examined to identify items with high outfit statistics (upper limit, 1.4). The analysis was repeated after item removal to confirm that all remaining items adhered to the Rasch model.

Factorial validity was tested by subjecting item measure residuals to PCA, and person and item measure separation reliabilities were obtained. To compare these outcomes with those of past analyses,[2] PCA with varimax rotation was applied to the raw scores of the 17 items.

TABLE 1. Item Fit Statistics from the DLTV-22, which Employs a Four-Point Ordinal Response Scale

Entry Number	Item	Logit Measure (Error)	Infit MNSQ	Outfit MNSQ
14	Reading normal size newsprint	2.13 (0.11)	0.99	0.80
7	Distinguishing a person's features across the street	1.87 (0.11)	0.83	0.70
16	Reading correspondence	1.55 (0.11)	1.09	0.91
22	Feel confident to walk around unfamiliar neighborhood	0.97 (0.10)	1.19	1.16
1	Distinguishing a person's features across the room	0.96 (0.10)	0.89	0.81
6	Reading road signs/street names	0.85 (0.10)	0.84	0.72
3	Watching television	0.46 (0.10)	0.59	0.65
17	Signing documents (e.g. checks)	0.38 (0.11)	1.06	1.04
19*	Difficulty adjusting to brightness after the dark	0.18 (0.11)	1.42	2.29
20*	Difficulty adjusting to darkness after light	0.01 (0.11)	1.65	2.52
18	Identifying money from a wallet	−0.02 (0.11)	0.74	0.64
12	Cutting fingernails	−0.19 (0.11)	1.01	0.84
15	Reading newspaper headlines	−0.33 (0.11)	1.03	0.81
4*	Seeing steps and using them	−0.56 (0.12)	0.88	1.81
10	Pouring a drink	−0.61 (0.12)	0.59	0.99
21	Feeling confident to walk around a familiar neighborhood	−0.66 (0.12)	1.02	0.92
9	Distinguishing a person's features at arm's length	−0.68 (0.12)	1.03	0.75
13	Using kitchen appliances	−0.71 (0.12)	0.91	0.68
5	Enjoying the scenery if out for a drive	−1.06 (0.13)	0.90	0.78
8	Recognizing seasonal changes in the garden	−1.29 (0.14)	1.00	0.75
11	Cutting up food on a plate	−1.43 (0.14)	0.70	0.62
2*	Noticing objects off to either side	−1.82 (0.16)	1.23	4.47

The hierarchy of items and fit statistics are shown. Column 3 shows the item measure (log odds of the probability), which is a reflection of the difficulty of performing the task and is used to order the items, with the hardest task at the top and the easiest at the bottom (figures in parentheses show the variation in logit measure). Columns 4 and 5 are infit and outfit mean squares, respectively.

* Taking the recommended upper limit of 1.4, these four items have outfit mean squares greater than 1.4, and of these, only one, "Difficulty adjusting to darkness after light," exhibits infit and outfit statistics above the limit.

TABLE 2. Item Fit Statistics from the DLTV-11 when Employing the Five- and Four-Point Ordinal Response Scales

Entry Number	Item	Logit Measure (error)		Infit MNSQ		Outfit MNSQ	
		5-Point	4-Point	5-Point	4-Point	5-Point	4-Point
8	Reading normal size newsprint	1.60 (0.07)	2.16 (0.10)	0.97	0.95	0.87	0.85
1	Distinguishing a person's features across the street	1.17 (0.07)	1.43 (0.09)	0.98	1.08	1.06	1.00
10	Reading correspondence	1.04 (0.07)	1.36 (0.09)	1.01	0.90	0.89	1.05
4	Watching television	0.42 (0.07)	0.50 (0.10)	0.92	0.84	1.06	1.01
11	Identifying money from a wallet	−0.02 (0.07)	−0.05 (0.10)	0.83	0.84	0.74	0.76
5*	Seeing steps and using them	−0.28 (0.08)	−0.40 (0.10)	1.40	1.21	1.67	1.44
9	Reading newspaper headlines	−0.51 (0.08)	−0.62 (0.11)	1.16	1.29	0.83	0.99
6†	Pouring a drink	−0.75 (0.08)	−0.91 (0.11)	0.74	1.18	0.89	1.56
3*	Noticing objects off to either side	−0.78 (0.08)	−0.97 (0.11)	1.30	0.75	1.77	0.88
2	Distinguishing a person's features at arms length	−0.86 (0.08)	−1.21 (0.12)	0.76	0.77	0.67	0.69
7	Cutting up food on a plate	−1.02 (0.09)	−1.28 (0.12)	0.83	0.94	0.89	1.05

The items are ordered with the hardest task at the top and the easiest at the bottom (numbers in parentheses show the variation in logit measure).

* When a five-point scale was used, only two items had outfit mean squares which fell outside the recommended limit of 1.4.

† When categories were collapsed to a four point scale, only one item had an outfit mean square that lay above the 1.4 limit. ("Seeing steps and using them" was just slightly above 1.4).

RESULTS

Participant Demographics

The responses from 186 participants constituted the first dataset. These participants had a mean age of 72.3 ± 6.4 (SD) years, 86 (43%) were men, and mean DVA in logMAR in the better eye was 0.45 ± 0.71 (SD).

Agreement between the 11 items of the DLTV-22 that constituted the reduced questionnaire (DLTV-11) and the remaining items of the DLTV-22 was excellent (Fig. 1).

Participants included in the second dataset (n = 324) had a mean age of 73.8 ± 10.3 years, 125 (39%) were men, and the mean DVA in the better eye was 0.39 ± 0.63 (SD) logMAR.

Rasch Analysis

Item measure statistics for the DLTV-22 are shown in Table 1 and that for the DLTV-11 when scored as a five-point scale and when collapsed to a four-point scale are shown in Table 2.

Assessment of the Rating Scale

Category frequency counts for the DLTV-22 (Table 3) show that the average measure is evenly separated: The largest counts are in the highest categories and the average measure values are close to the observed values for each of the categories—indicating that, overall, the rating scale is being used as predicted. The counts for the DLTV-11 (Table 4) did not increase monotonically and are unevenly separated. The expected measure, although close to the average measure for the two extreme categories (1 and 5), displayed slight deviation in the intermediate categories of the five-point scale. The probability characteristic curve that plots the person minus item measure against the probability of choosing a particular category is shown for the item "Reading newsprint" for the DLTV-22 and -11 in Figures 2 and 3, respectively. All other items in both instruments had probability characteristic curves that were identical with those shown in Figures 2 and 3.

Figure 2 (DLTV-22 using the four-point scale) shows that the peak for category 3 is higher than that for category 2, and the step calibrations are evenly separated. In Figure 3 (DLTV-11 with the five-point scale) the desired separation of modes is observed, but category 2 has a higher peak than either category 3 or 4. Category 3 has a very narrow range indicating that it is not used consistently and that it should be combined with category 4.[6–8] When categories 3 and 4 are combined (Table 5), step calibrations are more evenly separated. Reducing the number of response levels to three (by a further combination of categories 1 and 2) resulted in an additional increase in person separation (Table 6).

Scatterplots of the mean person raw score against person measures for the DLTV-22 (Fig. 4) and DLTV-11 (Fig. 5) indicate that the average rating changes monotonically, but not linearly, with the magnitude of the latent person trait. The curves were fitted to the data with the following equation:

$$\beta_n = a\ln\left(\frac{R_n}{(m-1)-R_n}\right) + b, \qquad (2)$$

TABLE 3. Rasch Output from Analysis of the DLTV-22

Category Label	Category Count	Average Measure	Expected Measure	Outfit MnSq	Step Calibration	Coherence	
						M→C	C→M
1	566	−1.42	−1.36	0.89	—	75%	44%
2	606	−0.17	−0.24	1.04	−0.85	35%	43%
3	919	0.90	0.87	1.48	−0.12	40%	52%
4	1789	2.57	2.59	1.01	0.97	82%	74%

Column 1 shows the four-step rating scale of the DLTV. The category counts increased monotonically and the average measure was evenly separated for each point on the scale. Column 6 which is the step calibration for each response, also illustrates even separation, confirming that the scoring of the DLTV-22 fits the Rasch model. Columns 7 and 8 are Guttman's coherence for each category. The Guttman coherence across all four responses was calculated and found to be 64%.

TABLE 4. Rasch Output from Analysis of the DLTV-11

Category Label	Category Count	Average Measure	Expected Measure	Outfit MnSq	Step Calibration	Coherence	
						M→C	C→M
1	317	−1.36	−1.41	1.21	—	68%	32%
2	454	−0.67	−0.57	0.92	−1.35	46%	52%
3	414	0.32	0.23	1.01	−0.08	32%	42%
4	668	1.17	1.15	1.00	0.19	38%	47%
5	1143	2.45	2.46	1.05	1.24	79%	66%

Column 1 shows the five-step rating scale of the DLTV-11. The category counts did not increase monotonically. The step calibration between the third and fourth point was small and uneven, suggesting that category 3 is a misfit. Columns 7 and 8 are Guttman's coherence. The coherence across all categories was calculated and found to be 57%. Categories 3 and 4 had the least amount of coherence, implying that they should be combined. These findings show that the five-point scoring system of the DLTV does not fit the Rasch model.

where β_n is the estimated person measure, R_n is the average raw score for person n across all items, m is the number of categories in a given scale, and a and b are linear regression coefficients estimated from log average raw score against person measure.

Content Validity

Person-item maps of group 1 (those with AMD in both eyes and whose VA is worse than 0.3 logMAR in the better-seeing eye), demonstrates a normal distribution without floor or ceiling effects (Fig. 6). A similar person-item map for group 2 (those with AMD in one eye and whose VA is $<$ 0.3 logMAR in the better-seeing eye) demonstrates a wide distribution, but there is evidence of skewing indicating clustering of participants at the top of the scale (Fig. 7).

Item Reduction

Three items in the DLTV-22 ("Difficulty adjusting to brightness after being in the dark," "Difficulty adjusting to darkness after being in the light," and "Noticing objects off to either side") had outfit statistics in excess of 2.0 and one item ("Seeing steps and using them") had an outfit mean square of 1.81. Applying an optimum upper limit of 1.4 as a cutoff resulted in the removal of these four items.[9] Repeating Rasch analysis on the 18 remaining items resulted in one further item "Pouring yourself a drink" continuing to show a high outfit mean square (2.5). After removal of this item, all remaining items had outfit mean squares within the optimal range. This adjustment also resulted in the decrease of the outfit mean square for response category 3 from 1.48 to 1.02, suggesting that there was an overrepresentation of easy items to begin with. PCA of person and item measure residuals revealed two domains with high person (0.86 and 0.77 for domains 1 and 2, respectively) and item reliabilities (0.99 and 0.98 for domains 1 and 2, respectively). On comparing the factor loadings from Rasch analysis for each domain with the factor coefficients obtained by PCA with varimax rotation, the outputs from the two methods were highly similar (Table 7).

DISCUSSION

The DLTV has previously been shown to be a useful instrument in assessing the impact of visual impairment in older adults.[2] In the present study, we extended our assessments of the DLTV

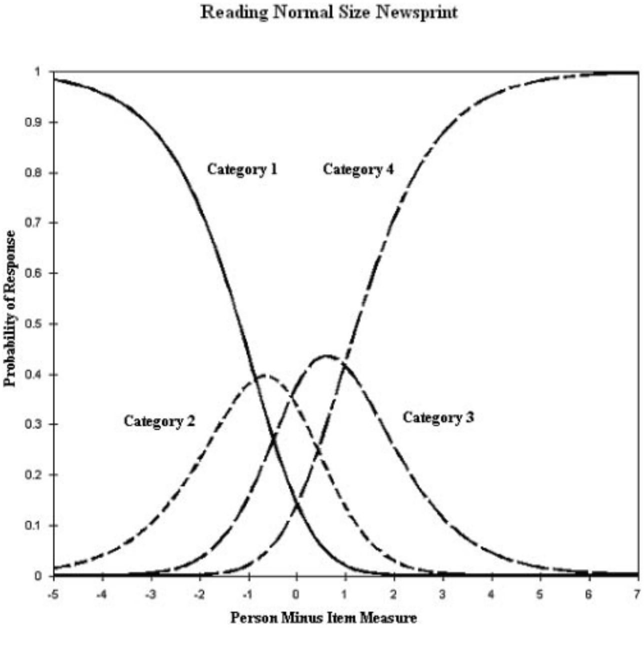

FIGURE 2. Probability characteristic curve for item 14 on DLTV-22 (Hardest Task). The person measure minus the item measure is plotted along the *x*-axis against the probability of response on the *y*-axis. The point at which each curve intersects another curve is known as the step calibration, the exact values of which can be found in Table 4. This is also the point at which a respondent is likely to choose a category over the previous category.

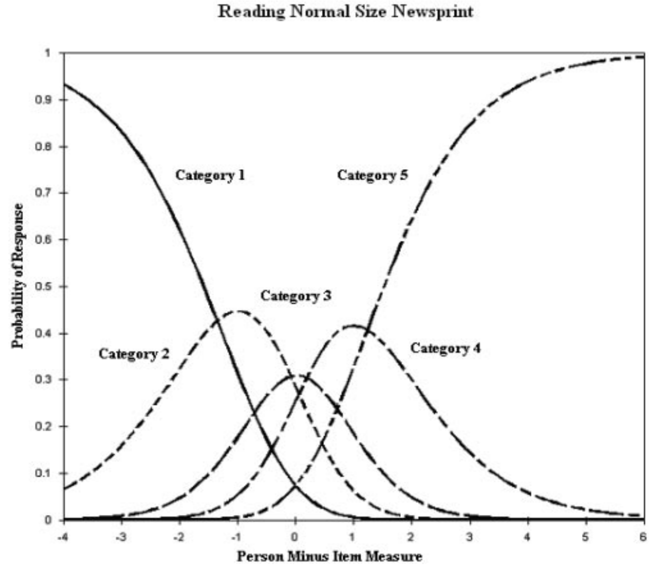

FIGURE 3. Probability characteristic curve for item 8 on the DLTV-11 (Hardest Task). The person measure minus the item measure is plotted along the *x*-axis against the probability of response on the *y*-axis. The point at which each curve intersects another curve is the step calibration, the exact values of which are shown in Table 5.

TABLE 5. Rasch Output from Analysis of the DLTV-11 with Categories 3 and 4 Combined

Category Label	Category Count	Average Measure	Expected Measure	Outfit MnSq	Step Calibration	Coherence	
						M→C	C→M
1	317	−1.69	−1.80	1.25	—	70%	40%
2	454	−0.65	−0.44	0.80	−1.48	46%	55%
3	1082	1.17	1.10	1.06	−0.57	59%	66%
4	1143	3.20	3.22	1.01	2.05	78%	72%

There was even separation through the four categories, and Guttman coherence, when calculated, improved to 65%. Results were then similar to the DLTV-22 and indicated the validity of combining categories 3 and 4.

to include the original version that uses a four-point scale and data obtained with a subset of 11 items with a response scale increased to five steps. We confirmed that the DLTV-11 and the remaining items of the overall DLTV-22 were measuring the same construct. We then sought to determine the scale that represented the more optimal structure, obtain a linear scoring system, and refine internal consistency. We used Rasch analysis to achieve these objectives. Rasch analysis is particularly useful in providing true linear scoring and deciding the appropriateness of the scoring scale. It combines the measure of the ability of the person with the measure of the difficulty of the item in question by subtracting one from the other. The observations in higher categories must be produced by higher measures, indicating that the average measures by category must be evenly separated throughout the rating scale.

Rasch analysis demonstrated that the four-point ordinal scale of the DLTV-22 was clearly optimal in comparison with the five-point scale. The intervals between categories were uneven (shown in Fig. 3), with the least separation between categories 3 and 4 suggesting that these should be collapsed into one. Seeing the results of this adjustment, we can conclude that the categories may be used interchangeably.

The Rasch analysis separation values (Table 6) also showed that the five-point scale was more imprecise than the four-point scale, with the implication that the extra response level distorts the sensitivity of the instrument rather than improving it. In this context, our findings support the observation of Stelmack et al.[8] who used Rasch analysis on the VA LV VFQ-48 and observed that the five-point scale was suboptimal. In this instrument, reducing the five-point scale to four-point through the combination of categories 2 and 3 also yielded a better fit. The same investigators tested the validity of their approach by analyzing a further dataset of the VA LV VFQ-48 which used a four-point scale that revealed that Guttman's coherence was identical with that obtained after combination of categories 2 and 3.

TABLE 6. Response Scales, Person Separation Value, and Person Reliability for Each Category for the DLTV-22 and the DLTV-11

Questionnaire	Scale	Real Person Separation	Person Reliability
DLTV-22	4	2.95	0.90
DLTV-11	5	2.18	0.83
DLTV-11	4	2.40	0.85
DLTV-11	3	2.56	0.87

For the DLTV-11, the response scale was reduced to four by combining categories 3 and 4 and then to three by further combining categories 1 and 2. Column 2 shows the response scale used for each questionnaire. Column 3, which is the person separation, shows that maximum separation occurred with the DLTV-22. For the DLTV-11, person separation was least with the five-point scale but increased as the response scale decreased. Column 4 is the person reliability and shows the maximum separation to be with the DLTV-22.

Other investigators have demonstrated similar findings on other visual function instruments. Pesudovs et al.[4] found that the five-point scale of the ADVS (Activities of Daily Vision Scale) was suboptimal. Velozo et al.[10] who examined the VF-14 (another visual function instrument), detected underutilization of the lowest response categories and recommended that its five-point scale be reduced to three. In addition Velozo et al. report that the person separation increased from 2.37 with a five-point response scale to 2.53 with a three-point scale. In the present study in an AMD population, in which the frequency of bilateral visual loss is high, we determined person separation to be 2.95 (person reliability = 0.90) for the DLTV-22 (four-point scale) and 2.18 (person reliability = 0.83) for the DLTV-11 (five-point scale). Combining two of the response categories of the DLTV-11 to restore a four-point scale improved person separation to 2.40 (person reliability = 0.85) indicating improved instrument performance. However, the person separation was not as clear as that seen when the instrument was administered as a four-point scale and emphasizes the importance of having the optimal scale in place at the time of administration. Although a three-point scale appears to yield an even better fit of the Rasch model it is possible that the DLTV becomes less sensitive to changes in visual function, and therefore additional studies are needed to investigate this possibility.

An average raw score for each person across all items is computed on the basis of the Likert scoring system. The Rasch rating scale transforms this into an interval scale with the result of each person being given a person measure.[6,7] However, on plotting the average person raw score against the person measure we established that they are not linearly related for the

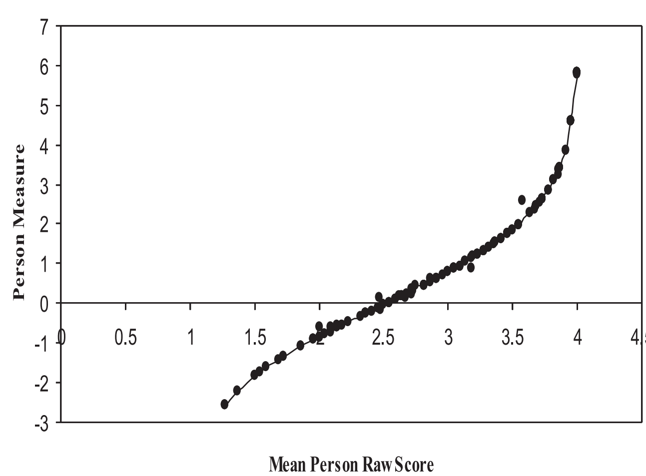

FIGURE 4. Scatterplot of the average mean person raw score against the person measures for DLTV-22. The average mean raw score for each person is plotted on the *x*-axis. The person measure estimated by Rasch analysis is plotted on the *y*-axis. The outliers in the scatterplot occur as a consequence of unexpected responses to DLTV-22 items. Regression coefficients, $a = 1.127$ and $b = -0.013$.

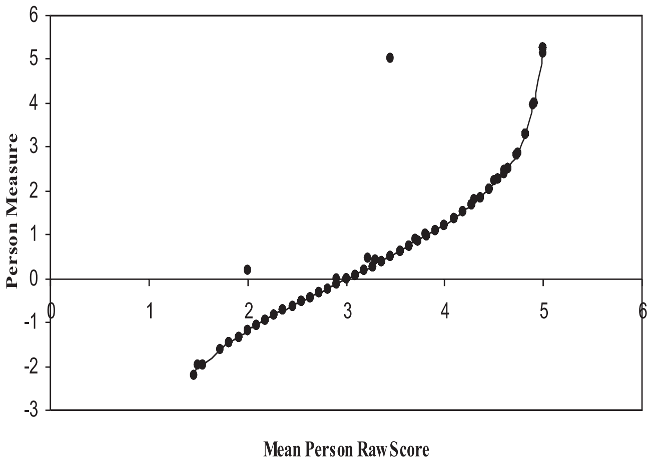

FIGURE 5. Scatterplot of the average mean person raw score against the person measures for DLTV-11. The average mean raw score for each person is plotted on the *x*-axis. The person measure estimated by Rasch analysis is plotted on the *y*-axis. The outliers in the scatterplot occur as a consequence of unexpected responses to DLTV-11 items. Regression coefficients, $a = 1.073$ and $b = 0.031$.

DLTV. This finding strongly supports the recommendations of Massof[3] who questioned the use of the Likert scale in analyzing VFQ.[11] We went on to find that it requires the generation of a double asymptotic nonlinear regression to transform the average person raw score to a Rasch person measure, and this process allowed us to use the DLTV-22 with a Rasch-adjusted scale without having to perform Rasch analysis.[11]

```
      <rare>|<more>
 4          +
            |  248
            |
            |  103
            |
           T|  214
 3          +  211  233  308  322
        X   |
            |  359
            |T 161  204
        X   |
        X   |  206  218  243
 2          +  332
           S|  220  229  302  350
            |  138  221  337
            |  203  230  235  321  324  347  353
            |S 156  238  245  355
        X   |  114
 1          +  113  167  318  327  363  368
       XX   |  234  342  361
            |  123  134  303  351  364  370
       XX M|   120  346  348  356  367
            |  115  219  330
            |  106  131  137  144  240  307  336  338  357  358
 0      X   +M 112  149  174  179  227  326  345
       XX   |  124  129  177  205  217
       XX   |  122  145  180  184  310
            |  228
      XXX   |  101  104  168  226  313  314  323
        X   |  111  178  328
-1      X S+   125  142  158  183  304
        X   |  109  181
            |S
       XX   |  117  130  182
            |
            |  157  212
-2          +  136  155
        X   |
           T|  102
            |T
            |  107  176
            |
-3          +
     <frequ>|<less>
```

FIGURE 6. Person–item map for DLTV-22 (group 1: both eyes affected by late AMD). Person-item map for the group of participants with DVA ≥ 0.3 logMAR in the better-seeing eye when both eyes are affected by the disease. The X's on the left represent the items in the DLTV-22 but do not define which item they represent. However, they do decrease in level of difficulty from top to bottom. Numbers on the right are the participants in the study. The distribution of patients and items are measured on the same scale.

```
        <rare>|<more>

 5          +  118  121  152  169  215  239  309  317  319
            |
            |
            |  108  159  160  171  222  224  237  246  301  312  315  333
               335  365
            |
            |
 4          +
           S|
            |  143  242
            |
            |  126  225  232  236  249  334  352
            |  241
 3          +  208  210
            |
            |  170  316
            |  209  216  366
           M|  116  146  325  354  362
            |
 2          +  141  247  369
            |  151
            |T 133  147  150  162  244  339
            |  202  349
       XX   |  105  213
        X   |  223  341  343
 1          +  110  320  329  360
       XX S|S  135
      XXX   |  140  340
            |  165
       XX   |
        X   |  166  344
 0          +M 128  139  148
       XX   |  231
        X   |
        X   |
        X T|   331
       XX   |S 132
-1     XX   +
        X   |
            |
            |
        X   |T
            |
-2          +
     <frequ>|<less>
```

FIGURE 7. Person–item map for DLTV-22 (group 2: only one eye affected by late AMD). Person-item map for participants with good VA<0.3 logMAR in the better-seeing eye when only one eye is affected by AMD. *Left* of the *dashed line*: the eleven items marked by an X; *right*: each participant.

We tested content validity by generating person–item maps for the DLTV-22. As visual function can be excellent even if acuity is poor in one eye, we dichotomized the second dataset into two groups: group 1, who have AMD in both eyes but with VA ≥ 0.3 logMAR in the better-seeing eye, and group 2, who have AMD in one eye but with VA< 0.3 logMAR in the better-seeing eye. The person–item map for group 1 (Fig. 6) illustrates the absence of floor or ceiling effects, confirming that the items of the DLTV-22 were well targeted. Group 2 (Fig. 7) in comparison exhibits ceiling effects for most items, indicating that the DLTV-22 is unlikely to discriminate between persons with excellent and good visual function. We further refined the content validity of the DLTV-22 through examination of infit and outfit statistics and the spacing between the items in relation to the distribution of persons in the model.[11] As mean square statistics have been defined such that the uniform value of randomness is indicated by 1.0, Rasch analysis indicates whether it would be prudent to remove items or persons that do not fit the model, thus allowing an improved measurement without losing information.[6,7,12,13] With this in mind, items with high outfit mean squares (Table 1) were removed, reducing the DLTV from 22 to 17 items. The outfit mean square for response category 3 was reduced to 1.20, leading us to conclude that the high outfit mean square observed in this category was due in part at least to items that were perceived as too easy.

In our previous analyses, we used PCA on raw scores to assign items to domains and identified the presence of four domains in the DLTV-22.[2,14] When the five items with high outfit Rasch statistics were excluded, the 17 remaining were assigned to two domains. These items segregated into the two domains in a fashion almost identical with the domain structure that we had established, with two minor differences (see

Table 7. Factor Coefficients from Principal Component Analysis on the Reduced 17-Item Questionnaire

DLTV with 17 Items	Rasch Coefficients (Sorted by Factor Loading)		Raw Scores	
	Domain 1	Domain 2	Domain 1	Domain 2
Distinguishing a person's features across the room	**0.21**		**0.723**	0.405
Watching television		**−0.05**	**0.692**	0.516
Enjoying the scenery if out for a drive		**−0.41**	0.219	**0.767**
Reading road signs/street names	**0.01**		**0.705**	0.458
Distinguishing a person's features across the street	**0.28**		**0.836**	0.291
Recognizing seasonal changes in the garden		**−0.40**	0.193	**0.776**
Distinguishing a person's features at arm's length		**−0.04**	0.376	**0.690**
Cutting up food on a plate		**−0.30**	0.221	**0.826**
Cutting fingernails		**−0.24**	0.406	**0.659**
Using kitchen appliances		**−0.35**	0.334	**0.763**
Reading normal-sized newsprint	**0.68**		**0.889**	0.182
Reading newspaper headlines	**0.34**		**0.567**	0.556
Reading correspondence	**0.72**		**0.906**	0.198
Signing documents (e.g., checks)	**0.34**		**0.751**	0.317
Identifying money from a wallet		**−0.10**	**0.598**	0.536
Feeling confident to walk around a familiar neighborhood		**−0.49**	0.353	**0.659**
Feeling confident to walk around an unfamiliar neighborhood		**−0.46**	**0.568**	0.484

The second and third columns give the factor loadings from Rasch analysis. The fourth and fifth columns give the factor coefficients when PCA (with varimax rotation) was applied. The items are highlighted in bold for the domains in which they fall.

Table 7). First, "Reading newspaper headlines," which was previously assigned to domain 2 moved to domain 1, and the item, "Do you feel confident to walk around your own neighborhood?," which was previously assigned to domain 3, was reallocated to the new domain 2. We believe it important to retain a domain structure, as previous studies have shown that domain 1 is sensitive to changes at the better extreme of the visual acuity scale, whereas domain 2 is sensitive to changes in the moderate range.[14] With respect to the remaining five items of the DLTV which did not fit into either domains 1 or 2, we would recommend analyzing them as individual items if the full DLTV-22 is used. The implications of removal of these five items has not yet been tested fully and it is possible that they may be of value in assessing disease states other than AMD or when applied to younger populations. We also believe that longitudinal studies are needed, to establish the sensitivity of the two domains to change in visual function.

With the establishment of the accuracy and precision of the measurement scale and with its present structure comprising 17 items within two domains, we contend that the DLTV constitutes an easily administrable, robust instrument for the assessment of self-reported functioning in patients with central visual loss. We confirm that a four-point ordinal scale constitutes an optimal scoring system for patients with AMD and that increasing the number of categories to five causes problems with assumptions of linearity. In summary, the use of Rasch analysis has confirmed the appropriateness of the rating scale, improved content validity and provided a linear scoring system for the DLTV.

Acknowledgments

The authors thank the SFRAD Study Group: Alan Bird, Ian Chisholm, and Gilbert Mackenzie.

References

1. Hart PM, Chakravarthy U, Stevenson MR, et al. A Vision specific functional index for use in patients with age related macular degeneration. *Br J Ophthalmol*. 1999;83:1115–1120.
2. Hart PM, Stevenson MR, Montgomery A-M, et al. Further validation of the Daily Living Tasks Dependent on Vision (DLTV): identification of domains. *Br J Ophthamol*. 2005;89:1127–1130.
3. Massof RW. Likert and Guttman scaling of visual function rating scale questionnaires. *Ophthalmic Epidemiol.* 2004;11:381–399.
4. Pesudovs K, Garamendi E, Keeves JP, et al. The Activities of Daily Vision Scale for cataract surgery outcomes: re-evaluating validity with Rasch analysis. *Invest Ophthalmol Vis Sci*. 2003;44:2892–2899.
5. Hart PM, Chakravarthy U, MacKenzie G, et al. Visual outcomes in the Subfoveal Radiotherapy Study: a randomised controlled trial of teletherapy for Age-Related Macular Degeneration. *Arch Ophthalmol*. 2002;120:1029–1038.
6. Wright BD, Mok MMC. An overview of the family of Rasch measurement models. In: Smith EV, Smith RM, eds. *An Introduction to Rasch Measurement: Theory, Models and Applications*. Maple Grove, MN: JAM Press; 2004:Chap 1.
7. Linacre JM. Optimizing rating scale category effectiveness. In: Smith EV, Smith RM, eds. *An Introduction to Rasch Measurement: Theory, Models and Applications*. Maple Grove, MN: JAM Press; 2004:Chap 11.
8. Stelmack JA, Szlyk JP, Stelmack TR, et al. Psychometric properties of the Veterans Affairs Low Vision Visual function Questionnaire. *Invest Ophthalmol Vis Sci*. 2004;45:3919–3928.
9. Bond TG, Fox CM. The question of model fit. *Applying the Rasch Model-Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates; 2001:Chap 12.
10. Velozo CA, Lai JS, Mallinson T, et al. Maintaining instrument quality while reducing items: application of Rasch analysis to a self-report of visual function. *J Outcome Meas*. 2000;4:667–680.
11. Massof RW. Application of stochastic measurement models to visual function rating scale questionnaires. *Ophthalmic Epidemiol*. 2005;12:103–124.
12. Massof RW, Fletcher DC. Evaluation of the NEI visual function questionnaire as an Interval Measure of Visual Ability in Low Vision. *Vision Res*. 2001;41:397–413.
13. Smith HJ, Dickinson CM, Cacho I, et al. A randomized controlled trial to determine the effectiveness of prism spectacles for patients with age-related macular degeneration. *Arch Ophthalmol*. 2005;123:1042–1050.
14. Stevenson MR, Hart PM, Montgomery AM, et al. Reduced vision in older adults with age-related macular degeneration interferes with ability to care for self and impairs role as carer. *Br J Ophthalmol*. 2004;88:1125–1130.