Article

Optimizing the ULV-VFQ for Clinical Use Through Item Set Reduction: Psychometric Properties and Trade-Offs

Gislin Dagnelie[1], Pamela E. Jeter[1], and Olukemi Adeyemo[1]; PLoVR Study Group[1,*]

[1] Lions Vision Research & Rehabilitation Center, Department of Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Purpose: We examine the dimensionality of the 150-item visual functioning questionnaire for individuals with ultralow vision (ULV-VFQ) and develop representative abbreviated versions, facilitating clinical use, while retaining compatibility with a 17-item performance assessment.

Methods: Subsets with 50 and 23 items covering the full difficulty range were selected, with evenly spaced item measures (IMs) and good representation of visual aspects and functional domains. Person measures (PMs) for the anchored subsets were derived through Rasch analysis of data from 80 respondents.

Results: Fit statistics for the reduced item sets were similar to those for the full set, with reliabilities at or above 95%. Mean PMs in the reduced sets were within 0.8 standard errors (SEs) of those in the full set. SEs of the PMs increased from the SE for 150 items, roughly in inverse proportion with the square root of the set size. Unexplained variance levels (24%–27%) and variance of the first unexplained factor (3.3%–3.9%) were close to those (30% and 2.6%) for 150 items. Differential item functions for omitted items were negligible. Aspects and domains are adequately represented in the reduced sets.

Conclusions: Self-reported visual ability can be measured accurately using appropriately chosen anchored subsets of the ULV-VFQ. Functional ability of individuals with ULV is characterized adequately by a single dimension.

Translational Relevance: The ULV-VFQ50 and ULV-VFQ23, using anchored IMs from the 150-item ULV-VFQ, provide an efficient and reliable self-report assessment of visual ability in individuals whose visual impairment is too severe for assessment with VFQs currently in use.

Introduction

Over the last two decades the use of calibrated visual functioning questionnaires (VFQs) has gained acceptance as an important component of outcome measurement in clinical trials[1] as well as low vision rehabilitation.[2] Psychometric techniques, such as Rasch analysis,[3] convert rating scales (e.g., degree of difficulty reported by a given respondent for a particular visual activity) into interval scales. Thus, the VFQ has become a measurement tool that assigns an ability score (Person Measure; PM) to each respondent, and a difficulty score (Item Measure; IM) to each question. This has allowed researchers, clinicians, and rehab workers to draw quantitative and statistically supported conclusions from patient-reported effects of treatments whose impact might be difficult to quantify otherwise.

Traditional VFQs require the respondents to have sufficient visual ability to perform daily activities shared by those with normal vision, albeit with the assistance of optical and opto-electronic aids; but these VFQs are of limited use in populations with profound vision loss. If the patient population under study has ultralow vision (ULV[4]; defined as vision allowing perception of light, projection, and/or movement, but no or very limited form vision) at baseline, the use of patient-reported outcomes becomes problematic. Recently, the use of VFQs has been expanded to populations with very low and ultralow vision through the introduction of specifically constructed instruments, the Impact of Visual

Impairment in Very Low Vision (IVI-VLV[5]) and the ultralow vision visual functioning questionnaire (ULV-VFQ[6]). As explained by Jeter et al., an important difference between those instruments is the quantity that is being measured. The IVI-VLV measures impact, that is, a combination of difficulty, importance, and emotional effects of severe vision loss, while the ULV-VFQ measures visual ability of the respondent. Another difference is the target population: The items of the IVI-VLV require more vision than those of the ULV-VFQ. During the development of the ULV-VFQ we demonstrated that even those without functional form vision can make consistent judgments of the visual ability required to perform activities that have very low visual demand. The items for the ULV-VFQ were derived from a much larger set of activities reported by 45 individuals with ULV in 6 focus groups.[7] The common thread for these activities was that they can be performed with very limited vision, but not without vision. ULV-VFQ items selected from these activities shared the property that they would be sufficiently familiar to most respondents, and that their wording was unambiguous with respect to lighting, contrast, and other important aspects that would affect the difficulty of the task.

A drawback of the 150-item ULV-VFQ is that it is time-consuming. It does not lend itself to use in a busy clinical setting, nor in a clinical trial where a VFQ is only one of a large set of assessments performed during study visits, at baseline, and during follow-up. One or more shortened versions of the questionnaire would be needed for use in such settings. This raises the question whether items can be eliminated from the ULV-VFQ without appreciably altering its psychometric properties. Specifically we will need to examine the following aspects:

- The dimensionality of the latent trait(s) underlying the visual ability reflected in responses to the ULV-VFQ. It has been reported that visual functioning in normal and low vision falls along two dimensions, roughly characterized as identification using central vision and detection using peripheral vision.[8] A preliminary analysis of the 150-item ULV-VFQ responses suggested that 70% of all variance could be explained by a single variable, with the first unexplained dimension amounting to just 2.6% of total variance. It will be important to ascertain that shortened versions of the instrument retain similar properties.

- Seven separate visual aspects were represented in the items of the ULV-VFQ, with contrast (93/150), luminance (17/150), distance (11/150), and movement (8/150) as the most important ones. Familiarity, environmental light, size, and a variety of minor aspects governed the remaining 21 items. More than half of the items were dependent on multiple aspects.[7] Analysis of the IMs showed that these aspects are not evenly represented: size, distance, and familiarity are represented in items with greater difficulty within the ULV range, while items governed by luminance and environmental light cluster near the low end of the ability/difficulty scale. It will be important to have these aspects represented evenly in shortened versions.

- All 4 functional domains[8] were represented in the ULV-VFQ items, but in proportions very different from those governing normal vision: visual information gathering (VisInfo, 107/150), visually guided movement (EyeHand, 30/150), mobility (10/150), and "reading" (3/150). In the context of ULV, "reading" encompasses any residual shape perception that can be accomplished visually. Reading was represented by a few items near the top end of the ULV difficulty range, whereas the other domains were evenly represented across the range. Shortened versions of the ULV-VFQ will need to include all four domains.

- We have developed, on the basis of 17 suitable items in the ULV-VFQ, a set of standardized activities that can be performed by individuals with ULV, and are currently in the process of calibrating these activities (to be reported elsewhere). Shortened versions of the ULV-VFQ should include these 17 items to allow comparison between self-report and measured performance, in ULV rehabilitation and as part of future clinical trials.

- Finally, most importantly, we have determined the IMs, that is, difficulties, of all 150 items in the ULV-VFQ. As reported[6] these items span approximately 6 logits, with substantial redundancy in the central 2.5 logits, and sparse items outside this range. Shortened versions of the ULV-VFQ should retain the sparse items while eliminating redundant ones, to maintain a wide range of IMs, and, thus, allow assessment of a wide range of visual abilities.

We examined the IM and PM distribution, dimensionality, and representation of visual aspects and domains in the full ULV-VFQ and in reduced

item sets, and derived three instruments with 50, 23, and 17 items, respectively, that retain the most important aspects of the full ULV-VFQ.

Methods

Data Source: Instrument, Respondents, and Data Processing

To examine item dimensionality and the effects of item number reduction on the psychometric properties of the ULV-VFQ, we made use of the data set collected in the second calibration round of the 150-item ULV-VFQ.[6] Briefly summarized, the items were derived from a larger set of 760 activities for which focus groups of individuals with ULV reported still making use of their rudimentary vision.[7] A larger pool of individuals with ULV was recruited as respondents for the ULV-VFQ prototype. Table 1 lists the demographics and vision status of the respondents. Selection and wording of the items were refined, on the basis of Rasch analysis and subject feedback in round 1, and a second round administration followed. To analyze the four-level responses, ranging from 1 (impossible) to 4 (not difficult), by 80 individuals with ULV we used a Rasch-Andrich rating model in Winsteps (version 3.91.2; available at winsteps.com[9]), and estimated PMs and IMs, in logits.[10] PMs ranged from –5.13 to 4.36, IMs from –3.39 to 2.70 logits. The excellent psychometric properties of the data set – specifically, the high overall reliability, low PM/IM standard error (SE), and generally small infit estimates – made these data a suitable test set for the present development study. The excellent adherence of the data to the Rasch model (PM and IM reliability were 0.99 and 0.97, respectively) justifies the idea that subsets of the ULV-VFQ with lower item counts may have acceptable psychometric properties and, thus, be suitable for use in clinical settings.

Analytic Approach

The development and refinement of the ULV-VFQ was done as part of the Prosthetic Low Vision Rehabilitation (PLoVR) curriculum development program. The overarching goal of this program is to create a system of assessments and training tools that include self-report and scored visual performance measures. In parallel with the ULV-VFQ, we have created and are calibrating a set of performance measures similar to 17 of the ULV-VFQ items (to be

reported elsewhere). Therefore, these 17 items, out of the 150 items in the ULV-VFQ, were retained while creating shortened versions of the instruments. The remaining "free" items for the intermediate versions (i.e., 33 for the 50-item version and 6 of those 33 for the 23-item version) were selected to optimize 2 criteria: even distribution of the IMs and even representation of the visual aspects and domains. Response category thresholds and IMs were anchored while re-analyzing the data by subset size, visual aspect, or domain, to allow comparison of PM estimates and SEs. Scree plots were generated to examine general fit to the Rasch model and dimensionality of the data. PMs for the full data set and for individual aspects and domains were submitted to a Principal Factor Analysis in JMP (SAS Institute, Inc., Cary, NC) to examine differential loading across the first two factors. Differential person functioning was examined for visual aspects and domains as warranted by the results.

Specifically, we performed the following steps:

1. The 17 items corresponding to the set of performance activities were placed in all 3 subsets (50, 23, and 17 items) as required items.
2. After ordering the 150-item set by IM, a 50-item subset was created by choosing 33 additional "free" items, such that the resulting 50 items would be at roughly equidistant intervals. Alternate items with IMs near those of the 33 selected ones were chosen as possible substitutes.
3. Selected and alternate items, along with the 17 required items, were screened for representation of visual aspects and visual domains, and alternate items from under-represented aspects and domains were substituted for selected ones as needed, while maintaining regular IM intervals, as closely as possible.
4. The process under steps 2 and 3 was repeated to obtain a 23-item set from the 50-item set, with appropriate spacing across the full IM range. Effectively, this meant retaining 6 of the previously selected 33 items, along with the 17 required ones.
5. As was the case for the creation of the 50-item set, an effort was made to retain optimal representation of visual aspects and domains.
6. IMs and category boundaries were anchored to those of the 150-item set, and Winsteps analyses were run for the 3 reduced item sets, as well as for visual aspects and domains within the 150-item set.

Table 1. Demographics of the 80 Respondents Included in the Analysis of the Full ULV-VFQ Data Set[6] and in the Present MS

Total Sample, Round 2, $n = 80$	Range = 24–88
Age, mean $\pm$ SD	61 $\pm$ 14
Sex	
Female	35 (43)
Male	45 (57)
Administration	
Phone	56 (70)
Online	24 (30)
Usable vision	
Yes	59 (74)
No	21 (26)
Visual impairment	
<20/400 to >CF	13 (16)
CF to >LP	25 (31)
LP and worse	37 (46)
Retinal implants (LP)	5 (6)
Diagnosis	
Retinitis pigmentosa	52 (65)
Age-related macular degeneration	3 (4)
Other retinal dystrophy	7 (9)
Glaucoma	7 (9)
Other	9 (11)
Missing	2 (3)
Education	
High school	8 (10)
Some college or technical school	24 (30)
Completed college or technical school	18 (22)
Some graduate/professional school	7 (9)
Completed graduate/professional school	21 (26)
Did not respond	2 (3)
Race	
Caucasian	69 (86)
African-American	5 (6)
Hispanic/Latino	3 (4)
Asian	1 (1)
Other	2 (3)
Annual income	
$\leq$20,000	13 (16)
$21,000–$40,000	13 (16)
$41,000–$60,000	11 (14)
$61,000–$80,000	11 (14)
>$80,000	22 (26)
Did not respond	10 (13)

Table 1. Continued.

Total Sample, Round 2, $n = 80$	Range = 24–88
Employment	
Employed	22 (26)
Self-employed	9 (11)
Retired	35 (44)
Semi-retired	1 (1)
Disabled	1 (1)
Unemployed	4 (5)
Student	4 (5)
Did not respond	4 (5)
Relationship	
Single- never married	12 (15)
Living with partner-not married	2 (3)
Married	43 (54)
Divorced or separated	14 (18)
Widowed	4 (5)
Did not respond	5 (6)
Living situation	
Alone	24 (30)
With family or Friends	54 (68)
Did not respond	2 (3)
Activity	
Active	65 (81)
Sedentary	12 (15)
No or ambiguous response	3 (4)

This Table describes respondent characteristics for Round 2 ULV-VFQ development ($n = 80$). CF, counting fingers; LP, light perception. Numbers in parentheses are percentages.
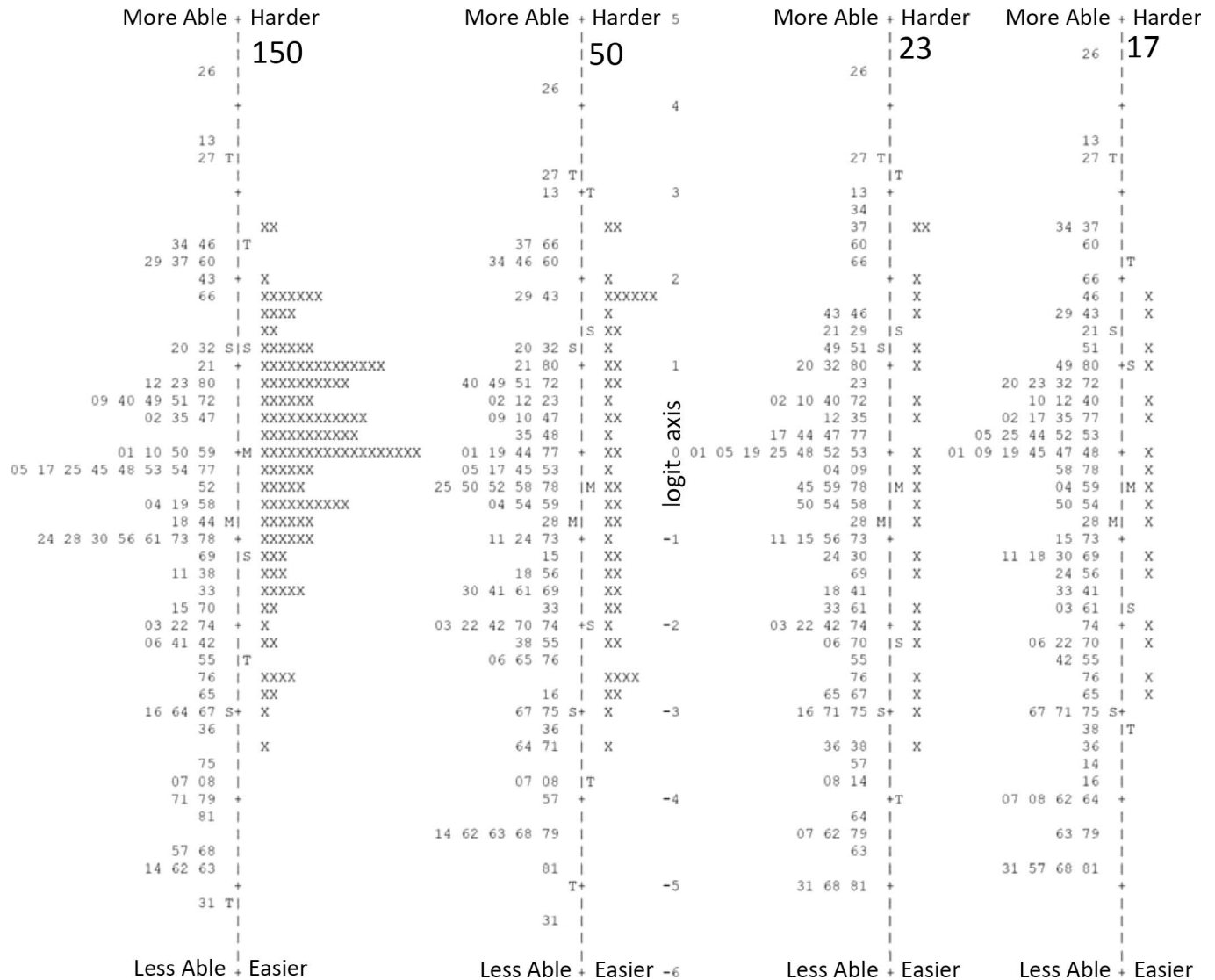
7. Scree plots were used to examine the magnitude and structure of unexplained variance.
8. PMs for the different sets were submitted to a Principal Components Analysis (PCA) to examine dimensionality of the underlying latent trait.
9. Differential PMs were computed for any subsets of items suggesting a difference in underlying latent traits.

Results

Subset Construction and Properties

As explained above, the items in the 150- and 17-item sets were fixed, and those selected for the 50- and 23-item versions were optimized for distribution along

Figure 1 (four Rasch person-item plots with headings "More Able — Harder" above and "Less Able — Easier" below, labeled 150, 50, 23, and 17; central "logit axis" with marks 5, 4, 3, 2, 1, 0, -1, -2, -3, -4, -5, -6)

Figure 1. Person (*leftward*) – Item (*rightward*) plots along a common logit axis, for the four ULV-VFQ versions, with item anchoring used in analyses for the reduced item sets. Item difficulty and person ability increase from *bottom* to *top*. Notice minor shifts and reduced resolution in PMs with decreasing item set size, especially for those with lowest visual ability.

the IM scale and for representation of visual aspects and domains. The four versions of the ULV-VFQ, along with the anchored IM estimates and SEs and assignments to visual aspects and functional domains, are included as Supplementary materials. Table 1 shows distribution measures for the IM estimates and intervals, PM estimates and SE, and assignment of items to primary visual aspects and functional domains. Figure 1 shows, for each of the 4 sets, the distributions of PMs (person IDs to left of the logit axis in each panel) and item measures (X symbols to the right of the logit axes). Each logit axis shows marks for the mean ("M") IM (0 by definition) and PM, and for one (S) and two (T) standard deviations.

The following properties of the distributions may be noted:

- IMs in all 4 panels are identical due item anchoring in the Rasch analysis. The only difference is that unused items are omitted before the analysis
- IMs for the 17 mandatory items ranged from −2.71 to 1.80 logits, that is, 74% of the 150-item range. The range for the 50- and 23-item versions was kept equal to the 150-item range.
- IM distributions for reduced item sets are centered lower than that for the full set. This is due to the skewed distribution of the original 150

translational vision science & technology

items – upward, that is, more demanding, items are overrepresented in the 150-item set and, therefore, have been preferentially removed.

- Many IMs in the 150-item set were statistically indistinguishable from those of neighboring items. Most of this redundancy was addressed in the reduction to 50 items, the remainder in the further reduction to 23 items.
- The 50-item set retains additional items at approximately −2.6 and +1.8 logits to make up for the sparseness of items near the top and bottom of the range.
- The overall PM reliability decreases with the number of items, but only to 0.95.
- As confirmed by the values in the PM block of Table 1, the PM distributions for smaller item sets are quite similar to that for 150 items. The median PM shifts by less than the SE of its own estimate.
- As expected, SEs on the PM estimates increase for smaller item sets, but no more than in inverse proportion with the square root of the set size.
- Visual aspects and domains were well represented in the 50-item version, and adequately in the 23-item version. The only reason for not retaining an aspect (50-item, size; 23-item, familiarity, other, and mobility) was the short distance between the IM and one of the 17 mandatory items, that is, items representing these aspects and domains were dropped in favor of retaining more even IM spacing.

Information, and Information-Weighted Item and Person Measure Fit Statistics
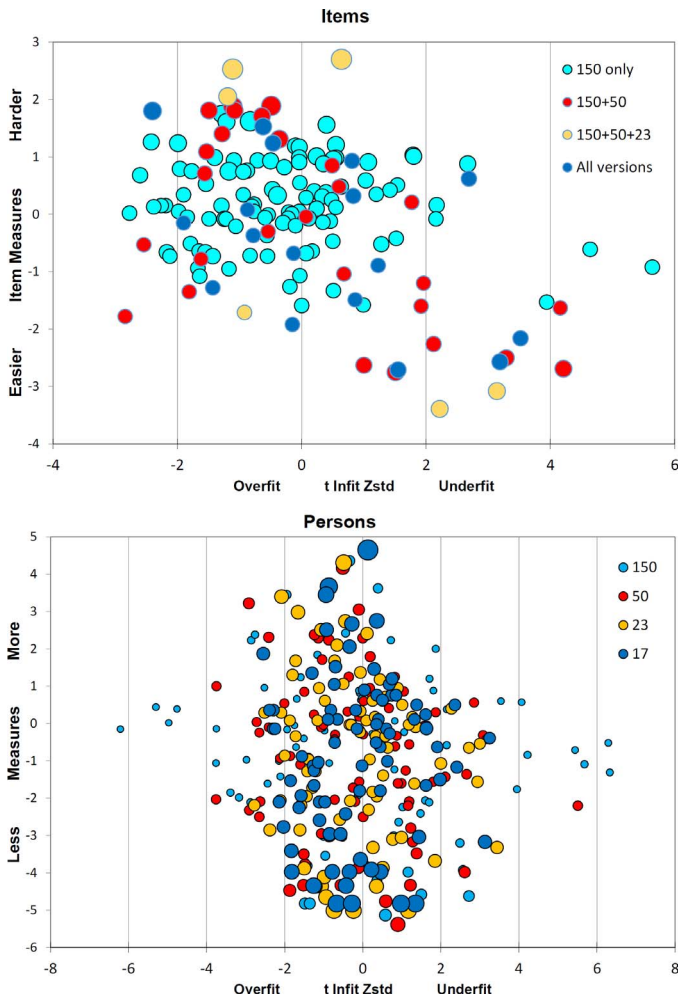
Each person submitting responses to a standardized VFQ contributes information to the Rasch analysis of the data set. Persons whose ability is well-matched by the items in the instrument will provide responses covering the full scoring category range, and, thus, contribute more information than persons who provide many floor or ceiling responses. Mathematically, this is reflected by smaller SEs for mid-range PM estimates than for those towards the ends of the range, as shown in the top panel of Figure 2. The fitted fourth order polynomial curves are very similar, showing a slight leftward shift of the minimum SE for sets with lower item counts; this is expected since these sets have lower median IMs (see Fig. 1).

Since the PMs in the 150-item response set are estimated from a larger number of items, one would expect their SEs to be smaller by a factor $\sqrt{(150/N)}$
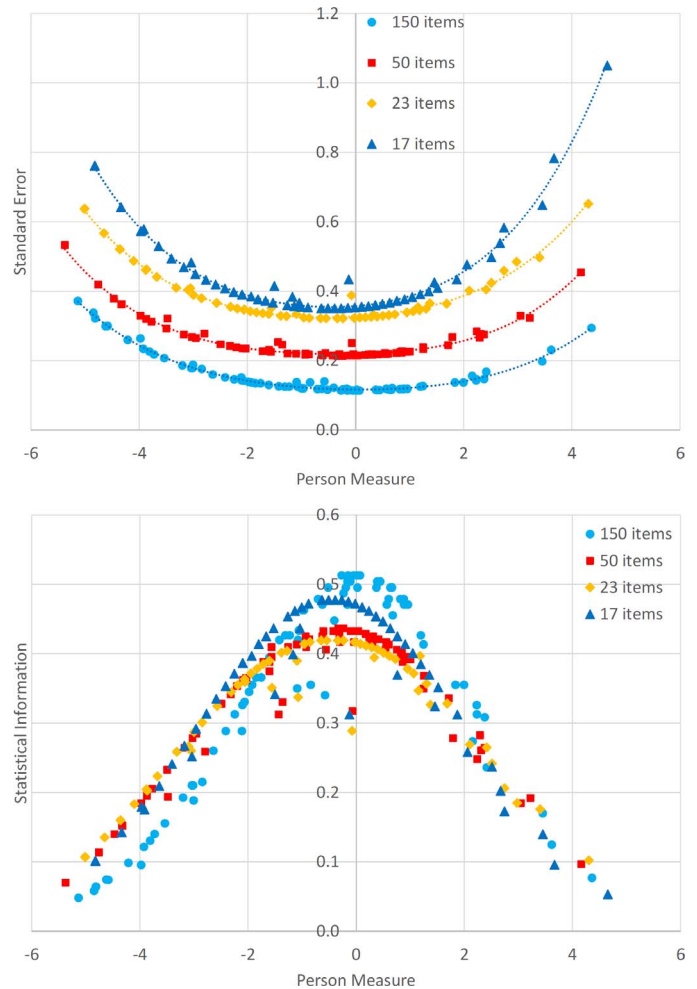
Figure 2. SE (*top*), statistical information ($[N \times SE^2]^{-1}$; *bottom*) gained, as a function of PM, for administration of 4 versions of the ULV-VFQ. The maximum amount of information gained is only slightly higher with 150 items than with lower item numbers. Information is gained even for PMs well outside the IM range (−3.39–2.70; −2.71–1.80 for the 17-item version), but SE rises sharply in this region, especially for 17 items. Notice the shift in the peak of the distribution for the smaller item sets: As the median item measure shifts, so does the person measure for which SE is smallest, and for which the highest information is gained.
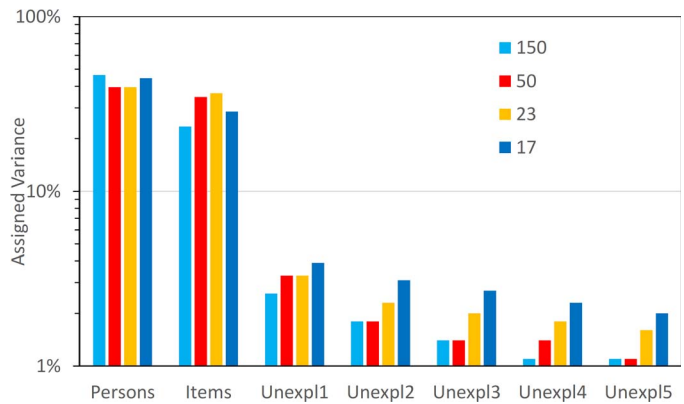
compared to those of the PM estimates for sets with a lower number of items. However, since the items are not regularly distributed, with more items in the 0 to +2 logit range, SE(150) should be lower in that range than for the smaller item sets, and higher elsewhere. This is confirmed in the bottom of Figure 2, where $SE \times \sqrt{N}$ has been plotted on a logarithmic scale. The curves are very similar, reflecting the near-constant value of the product of mean SE and $\sqrt{N}$ in Table 1; it is noteworthy that the smaller item sets yield noticeably lower normalized SEs for individuals

Figure 4. Scree plot of the explained (persons, items) and 5 largest unexplained variance components, for each of the 4 data sets, on a logarithmic scale. Total unexplained variance ranged from 30% (150 items) to 24% (23 items).

Figure 3. IMs and reliability Z-scores for the ULV-VFQ IMs (anchored; *top*) and PMs (unanchored; *bottom*). As indicated in the *top* legend, Items represented by *dark blue circles* are included in all four versions of the instrument, others only in the versions indicated by their markers. The 4 sets in the *bottom* do not coincide, since PMs were not anchored. Dot size in either panel is proportional to the SE of the Measure estimate, and, thus, lowest in the center of the vertical range (*both* panels) and for the largest item set (*bottom*).

in the −5 to −2 logit range, that is, those with the most severe vision loss. The curves for 50 and 23 items are smoothest, suggesting that the choice of items for those versions instruments leads to optimal consistency in the estimates.

The quality of the IM and PM fits can be examined through so-called bubble plots, in which the IM or PM is plotted against the Z-score of the information-weighted variance (infit Z-scores), which can be thought of as measurement variance that is influenced more by centered than by outlying PMs; bubble size is proportional to the SE of the estimate. High Z-scores in

bubble plots represent underfitted items (or persons), and $Z > 4$ generally is considered a poor fit. The top of Figure 3 shows the anchored item bubble plots for the 4 versions of the instruments, with color codes used to indicate in which version of the instrument each item was used. One may note that of the 5 poorly fitting items in the 150-item ULV-VFQ, only 2 remain in the 50-item version, and none in the smaller sets. This also is reflected in the bottom of Figure 3, where the PM bubble plots for the 4 data sets have been superimposed. Note that the bubbles for different sets do not coincide, since PMs were not anchored (unlike the items). Another noticeable effect of set size reduction is that there are fewer under- or overfitted PMs, but this is to be expected: Z-scores change inversely with the SE of the estimate, and this SE is larger for smaller item sets.

To examine the structure of the difference between the data and the Rasch model, and in particular whether this structure differs by set size, it is helpful to use the PCA on the residuals provided by Winsteps. The results of this analysis, in the form of a scree plot, are shown in Figure 4, with variance plotted on a logarithmic scale. Note that in all 4 sets at least 70% of the variance in the data is explained by the measures, that is, the Rasch model, and that the largest unexplained component is 4%, for the 17-item set, with subsequent components gradually decreasing. The 150- and 50-item sets show relatively large first unexplained components, but these represent only approximately 3% of the total variance. In other words, there is no appreciable structure to the unexplained variance. Unexplained components are larger for the smaller item sets, since these have far fewer components across which the unexplained

Table 2. Comparison of the 4 Versions (150, 50, 23, and 17 items) of the ULV-VFQ Discussed

Quantity	Measure	150 Items	50 Items	23 Items	17 Items
IM (logits)	Median	0.11	−0.34	−0.37	−0.37
	Min	−3.39	−3.39	−3.39	−2.71
	Max	2.70	2.70	2.70	1.80
IM spacing (logits)	Median	0.02	0.11	0.26	0.28
	Min	0.00	0.00	0.14	0.14
	Max	0.49	0.49	0.49	0.43
IM	Reliability	0.97	0.98	0.99	0.98
PM	Reliability	0.99	0.98	0.97	0.95
PM (logits)	Median	−0.67	−0.58	−0.60	−0.57
	Min	−5.13	−5.38	−5.02	−4.83
	Max	4.36	4.17	4.31	4.66
	Median Diff/SE	–	0.87	0.78	0.73
PM SE (logits)	Median	0.13	0.23	0.34	0.38
	Min	0.11	0.21	0.32	0.35
	Max	0.37	0.53	0.65	1.05
	Mean $\times \sqrt{N}$	1.94	1.82	1.83	1.83
Visual aspects (N)	Contrast	93	28	10	8
	Luminance	17	9	5	4
	Illumination	3	3	2	1
	Familiar	4	1	0	0
	Movement	8	3	4	4
	Distance	11	3	2	0
	Size	3	0	0	0
	Other	11	1	0	0
Visual domains (N)	Reading	3	1	1	0
	Mobility	10	3	0	0
	Visual motor	30	9	5	4
	Visual info	107	35	17	13

The top of the Table compares the (anchored) IMs and their spacing, in terms of median and range. The middle part shows shifts in the PM distribution as the item set is reduced. The bottom part compares the distribution of items across visual aspects and visual domains. The 150 items are those of the original ULV-VFQ[b]; the 17 items correspond to performance measures currently under development; the intermediate versions sought a compromise in terms of IM spacing and item visual aspects and domains.

variance is distributed: The dimensionality of the principal components is equal to the number of items.

Correlation, Visual Aspects, and Functional Visual Domains

As shown in Table 2, reducing the item set differentially removes items governed by different visual aspects or pertaining to different functional domains; for example, items governed by contrast and/or pertaining to VisInfo were removed in larger proportion than other items. It is important to know whether this affects the ability of the instrument to detect possible subscales or multiple dimensions of ULV. This can be examined by anchoring the IMs and performing separate analyses of responses to items grouped by aspect and by domain, and to then analyze the correlations and dimensionality of the resulting PMs.

The central columns of Table 3 show correlations among PMs for the subsets by aspect, in the top portion, and by domain, in the bottom portion. Numbers of items in each subgroup are given in

Table 3. Correlation matrices of Person Measures; top: Visual Aspects; bottom: Functional Domains. The central part in each half shows Pearson correlation coefficients for the person measures computed for all items (leftmost column) and for the item subsets governed by different Aspects or Domains. The rightmost two columns show, for the 50- and 23-item subsets, the Person Measure correlations with the full set, for the most common (Contrast) and all remaining Aspects, and for the most common (Visual Information) and all remaining Domains.

ASPECT	All(150)	C(93)	L(17)	I(3)	F(4)	M(8)	D(11)	S(3)	50 vs 150	23 vs 150
Contrast	0.996								0.984	0.956
Luminance	0.960	0.941								
Illumination	0.774	0.768	0.755							
Familiar	0.887	0.875	0.821	0.628						
Movement	0.941	0.929	0.885	0.728	0.851				0.978	0.962
Distance	0.941	0.925	0.895	0.692	0.865	0.899				
Size	0.894	0.882	0.854	0.657	0.833	0.826	0.866			
Other, 11	0.817	0.808	0.769	0.569	0.772	0.773	0.780	0.817		
DOMAIN	All(150)	R(3)	M(10)	VM(30)					50 vs 150	23 vs 150
Reading	0.817									
Mobility	0.953	0.751							0.968	0.944
Visual motor	0.985	0.797	0.931							
Visual info, 107	0.999	0.812	0.945	0.979					0.988	0.970

parentheses. Most correlations are well over 0.80; in fact the only ones under 0.80 are based on very few items in one or both groups, or where one group is the "other visual aspects" category. This suggests that disproportionate removal of items in one or more aspect or domain subsets should not have a major effect on the PM estimates.

The rightmost 2 columns in both portions of the Table address the question whether the elimination of a large proportion of certain items (e.g., for the reduction from 150 to 50 items: 65 Contrast vs. 35 non-Contrast; 72 VisInfo vs. 28 non-VisInfo) has an important effect on the PM estimate. Here, all correlations are over 0.97 for the 50-item subset, and over 0.94 for the 23-item subset, suggesting that the PM estimates for (grouped) aspects or domains are not substantially altered by set size reductions.

Similar correlations were calculated between PMs estimated from the 50 items and from the 100 original items that were omitted from the 50-item subset, and similarly for the 23 items and the 127 omitted items. These correlations also were quite high: The lowest value, 0.91, was found for the 6 non-VisInfo items in the 23-item set and the 37 non-VisInfo items excluded from that set. Note, however, that the mean SEs of the PMs for the 6 and 37 items were 0.79 and 0.35 logits, while the PM range was 9.3 logits; for such relatively noisy estimates a correlation of 0.91 can be considered quite good.

Dimensionality

Finally, even though we have shown that reduction of the item set can be done without appreciable loss of accuracy, and with no more than expected loss of precision in the estimates, it is important to examine whether the difference between visual aspects and/or functional domains may justify the introduction of subscales representing multiple dimensions in the latent trait (visual ability) assessed by the ULV-VFQ.

To do this, PM estimates for the aspect and domain subgroups were submitted to a PCA and Factor Analysis. Factor analysis showed that 84.0% of the variance in the correlation matrix for visual aspects could be explained by a single factor, and just 2.9% by the next largest factor. For the correlation matrix of PMs for functional domains these variance percentages were 90.7% and 5.6%. Figure 5 shows the projection of the PM estimates for item groups governed by different visual aspects (top) and domains (bottom) onto the first two Principal Components. Based on the vector maps in the left, it appears that a second dimension, differentiating illumination (environmental lighting) from familiarity, size, and distance for visual aspects, or reading
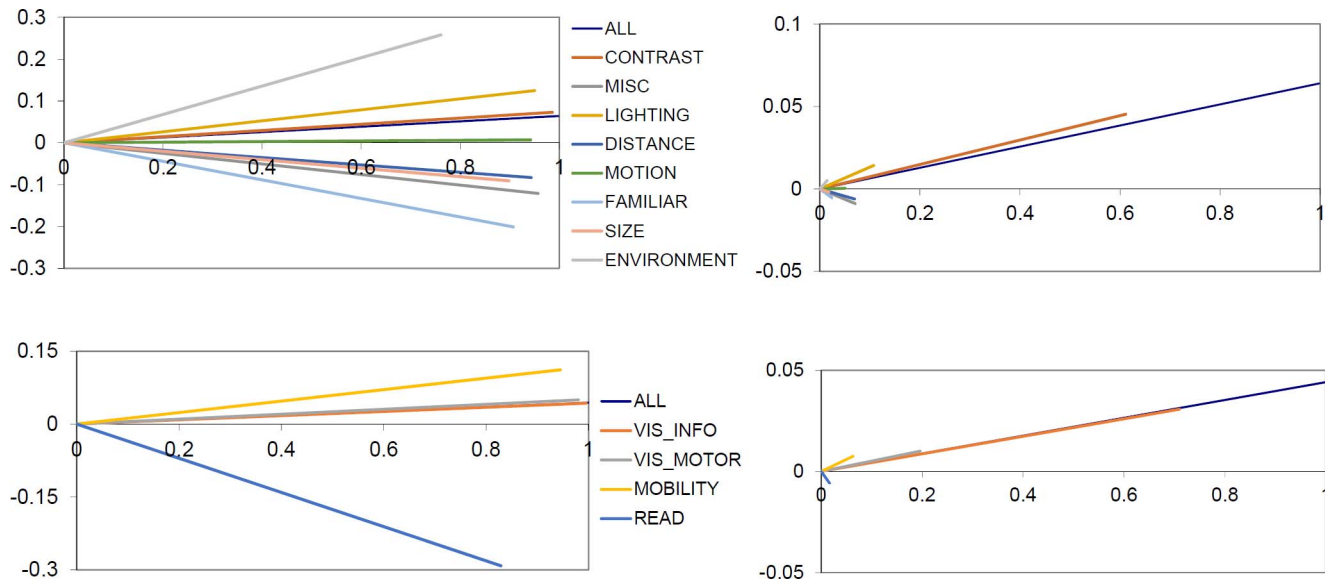
Figure 5. Person measure estimates for item groups governed by different visual aspects (*top*) and domains (*bottom*) projected onto the first two Principal Components derived from the PMs estimated for these item groups. The *left* shows projected vectors independent of item group size; the *right* shows the same vectors scaled according to item group size. Note the expanded vertical scale in the *right*.

from mobility for functional domains, may be significant. However, the percentage of the overall variance represented by this second dimension is only 1.8% for visual aspects, and 2.1% for functional domains. This is visualized in the right of Figure 5, where the vectors have been scaled according to the number of items they represent. Even with the expanded vertical scale it is clear that the distinction of a second dimension, and, therefore, of two subscales in the latent trait representation of ULV self-reports, is not justified by the data.

Discussion

The main goal of this study was to create shortened versions of the 150-item ULV-VFQ for clinical use that will provide, across a wide range of severely reduced visual ability, the most accurate PMs for a given item count as well as a good representation of items governed by different visual aspects and pertaining to different functional domains. An additional restriction was the retention of all 17 items corresponding to a concurrently developed set of performance measures. To accomplish this we retained evenly spaced items from the 150-item questionnaire while also preserving aspects and domains as much as possible. From the data presented above it appears that both the 50- and 23-item versions of the instrument meet the desired criteria, albeit that the shorter instrument will yield higher standard errors in

the PM estimates, that is, lower precision. The 17-item instrument, containing the items most suited to create a set of performance measures that can be administered in any setting, spans a shorter range, has less regular spacing, and large SEs especially at the end of the range, and for that reason is less well suited as an outcome measure for clinical studies.

From the fit statistics in Figure 3, it is clear that only 5 items were underfitted in the 150-item ULV-VFQ; only 2 of these remained in the 50-item version, and none on the smaller versions. Thus, at most 4% of the items are underfitted in the test population of 80 individuals with ULV. Administering the full questionnaire to additional ULV individuals may slightly increase the number of underfitted items, as can be seen from the change in person fit statistics with change in item set size: Smaller item sets appear to yield better results, but this is primarily due to the reduced set size, and, thus, a lesser degree of oversampling in the middle of the ability/difficulty range. This was confirmed by removing the 5 misfitting items from the 150-item set, which did not appreciably change the person fit. Thus, the PM misfits in the 150-item questionnaire are caused by item redundancy in the middle of the range, rather than by outlying or misfitting items.

One may wonder whether our results would have been different if we had run the analyses with a different data set from the one that was used to calibrate of the 150-item ULV-VFQ, or of the items had been

administered in a different order; a concern of item dependency was raised by one of the reviewers. We have no reason to suspect that this would be the case, as the 80 participants spanned a range in excess of the 150 items, and their data provided a consistent fit to the Rasch model. The concept of item dependency describes a situation in which not only the latent trait – in our case, functional reserve or visual ability – determines the ratings, but also extraneous factors, such as the order in which items are administered, or respondent fatigue or lack of interest. This is most often encountered as local item dependency, which expresses itself in unexpected relative IM shifts for items that should have similar IMs. In administering the ULV-VFQ we randomized the items and administered them in the same order to all respondents; thus, unless all respondents tired of the questions at the same rate, there is no reason to assume that item dependency had a role, and any such dependency would not be local. Thus, we feel confident that only minor adjustments of the item measures will be necessary as additional ULV individuals will contribute data in the future.

Respondents in our study were given an "opt-out" choice in addition to the 4 difficulty levels: If they felt that an item did not apply to them they could say so, and this answer was treated as missing data in the Rasch analysis. In the 150-item data set, the median number of items with this response was 2, and the mean 6.26, that is 4.2% of the number of items. With only 6 missing items on average, and even with the maximum number of 47 (31%) for one respondent in our population, the remaining redundancy was enough to get a precise PM. For the 50-, 23-, and 17-item versions a high number of missed items is obviously a concern. In our data sets the median numbers of missed items were all 0, and the mean numbers 1.3 (2.6%), 0.4 (1.7%), and 0.35 (2.1%); for the respondent with the highest opt-out rate, the numbers were 12 (24%), 6 (25%), and 5 (29%). Thus, if anything, our assessment of the reduced versions of the questionnaire was less affected by "opt-out" answers than the original 150-item analysis, and the missing item numbers were low enough not to have an appreciable effect on the resulting PMs for all but a few respondents. Thus, even with the "opt-out" choice available to respondents in a clinical setting, we expect few of them to give that answer to more than a few questions. If a high opt-out rate becomes a concern, the use of the adaptive version of the ULV-VFQ (Dagnelie G, et al. IOVS. 2015;56: ARVO E-Abstract 497) should be considered.

In constructing the 50- and 23-item version of the ULV-VFQ we were concerned that not all visual aspects and functional domains would be represented, in particular since we knew from our previous work[6] that certain visual aspects are differentially distributed along the visual demand axis. However, the finding that the data are essentially unidimensional greatly reduces that concern. Not only does it demonstrate that no meaningful subscales can be distinguished in the ULV-VFQ; it also shows that for the purpose of ability assessment, the choice among items governed by different visual aspects or pertaining to different functional domains is less important than the choice of items that evenly span the range of abilities to be assessed.

One may wonder if, rather than eliminating items on the basis of even spread, or visual aspect and domains, it would not be best to retain the most informative items. As shown in Figure 2, the most informative items are those in the center of the range, but that also is where most of the items are. Thus, if the sparseness of the information is taken into account, retaining items towards the end of the scale is important, particularly since the PMs of outlying (very able or disabled) individuals can only be estimated by virtue of outlying items.

We did not present results of a differential person functioning (DPF) analysis here. The reason for this is simple: Whether we grouped our data according to the item sets, functional domains, or visual aspects, in all cases the DPF values were very small. This is in line with the high correlations presented above, and with the lack of an appreciable second dimension in the PCA.

One remaining question may be whether there is any need for a calibrated 150-item instrument, if the shortened versions are so similar in properties. In our opinion, the merit of such an instrument are two-fold: (1) to continue building a large set of activities in the ULV range, as a calibration standard for future instruments, or (2) to provide an item bank for other ULV questionnaires, whether they are adaptive and select items most appropriate for each individual respondent, or have fixed item sets that concentrate on specific types of activities, or concentrate on certain visual aspects or domains.

In conclusion, we have derived from the 150-item ULV-VFQ two shorter ULV questionnaires for clinical use, the ULVVFQ-50 and the ULVVFQ-23, with anchored items, and have studied their psychometric properties; a third version, with 17 items, spans a shorter range and, therefore, is less appropriate in a population with a wide range of ULV. Which of the remaining instruments is preferred in a given setting

will depend on the willingness of the clinical investigators to spend more time administering the instrument, in exchange for greater precision of the PM estimates.

Acknowledgments

References

1. Hirneiss C, Reznicek L, Vogel M, Pesudovs K. The impact of structural and functional parameters in glaucoma patients on patient-reported visual functioning. *PloS One*. 2013;8: e80757.

2. Stelmack JA, Tang XC, Wei Y, Massof RW. The effectiveness of low-vision rehabilitation in 2 cohorts derived from the veterans affairs Low-Vision Intervention Trial. *Arch Ophthalmol*. 2012; 130: 1162–1168.

3. Massof RW. The measurement of vision disability. *Optom Vis Sci*. 2002;79: 516–552.

4. Geruschat DR, Dagnelie G. Restoration of vision following long-term blindness: considerations for providing rehabilitation. *J Vis Impair Blind*. 2016; 110: 5–13.

5. Finger RP, Tellis B, Crewe J, Keeffe JE, Ayton LN, Guymer RH. Developing the impact of Vision Impairment-Very Low Vision (IVI-VLV) questionnaire as part of the LoVADA protocol. *Invest Ophthalmol Vis Sci*. 2014;55: 6150–6158.

6. Jeter PE, Rozanski C, Massof R, et al. Development of the Ultra-Low Vision Visual Functioning Questionnaire (ULV-VFQ). *Trns Vis Sci Tech*. 2017;6(3): 11.

7. Adeyemo O, Jeter PE, Rozanski C, et al. Living with ultra-low vision: an inventory of self-reported visually guided activities by individuals with profound visual impairment. *Trans Vis Sci Tech*. 2017;6(3): 10.

8. Massof RW, Hsu CT, Baker FH, et al. Visual disability variables. II: The difficulty of tasks for a sample of low-vision patients. *Arch Phys Med Rehabil*. 2005;86: 954–967.

9. Linacre JM. *Winsteps* [computer software and user's guide]. Version 3.93.0. Beaverton, OR: Winsteps.com; 2017.

10. Fisher WP, Harvey RF, Kilgore KM. New developments in functional assessment: probabilistic models for gold standards. *NeuroRehabilitation*. 1995;5: 3–25.

translational vision science & technology