tvst

Article

Development of the Ultra-Low Vision Visual Functioning Questionnaire (ULV-VFQ)

Pamela E. Jeter, Collin Rozanski, Robert Massof, Olukemi Adeyemo, Gislin Dagnelie, and the PLOVR Study Group*

Lions Vision Center, Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Purpose: To develop and psychometrically evaluate a visual functioning questionnaire (VFQ) in an ultra-low vision (ULV) population.

Methods: Questionnaire items, based on visual activities self-reported by a ULV population, were categorized by functional visual domain (e.g., mobility) and visual aspect (e.g., contrast) to ensure a representative distribution. In Round 1, an initial set of 149 items was generated and administered to 90 participants with ULV (visual acuity [VA] $\leq$ 20/500; mean [SD] age 61 [15] years), including six patients with a retinal implant. Psychometric properties were evaluated through Rasch analysis and a revised set (150 items) was administered to 80 participants in Round 2.

Results: In Round 1, the person measure distribution (range, 8.6 logits) was centered at -1.50 logits relative to the item measures. In Round 2, the person measure distribution (range, 9.5 logits) was centered at -0.86 relative to the item mean. The reliability index in both rounds was 0.97 for Items and 0.99 for Persons. Infit analysis showed four underfit items in Round 1, five underfit items in Round 2 with a z-score greater than 4 cutoff. Principal component analysis on the residuals found 69.9% explained variance; the largest component in the unexplained variance was less than 3%.

Conclusions: The ULV-VFQ, developed with content generated from a ULV population, showed excellent psychometric properties as well as superior measurement validity in a ULV population.

Translational Relevance: The ULV-VFQ, part of the Prosthetic Low Vision Rehabilitation (PLoVR) development program, is a new VFQ developed for assessment of functional vision in ULV populations.

Introduction

Vision restoration therapies such as stem cell therapy,[1,2] gene therapy,[3–5] and other approaches,[6] including visual prostheses,[7–9] are currently being developed in research laboratories around the world. Among these promising possibilities, the Argus II Retinal Prosthesis System (Second Sight Medical Products, Inc., Sylmar, CA) was the first vision restoration treatment to become commercially available, receiving the European conformity mark in 2011 and Food and Drug Administration (FDA) approval for humanitarian use in the United States in 2013.[10] In practice, the level of vision achieved by these novel approaches is often reported as hand motions or worse and restricted to rudimentary form or shape perception.[11,12] When optotype visual acuity cannot be measured, clinicians customarily describe vision in terms of the patient's ability to see hand motions, light projection, light perception, or no light perception, but these classifications do not correspond to standardized measures. In ongoing clinical trials such as the Argus II feasibility study (ClinicalTrials.gov: NCT00407602), visual function has been assessed through ad-hoc measures,[7,13,14] because standardized visual function measures proved unable to quantify the visual outcome. These ad-hoc measures include minimal angle of resolution with high contrast gratings, square localization, direction of motion perception, and light sensitivity.[13,15,16] Importantly, these measures of visual function do not capture

functional vision,[17] that is, the ability to use vision in activities of daily living (ADLs),[18] and are uninformative regarding vision-related quality of life (VRQoL) of the individual.[19] Moreover, even for individuals with mild and moderate low vision[20–22] or with vision loss due to aging,[23–25] it has been shown that objective measures of vision do not correlate well with activities of daily living or psychological well-being. Therefore, to supplement the limited measures of visual function, there has been an increasing demand for patient-reported outcomes (PROs) that capture functional vision and VRQoL, especially in populations with severe vision loss.[26–30]

PROs reflect the patient's perceived changes in QoL and visual ability, and thus are central to rehabilitative outcomes measurement. Massof[31,32] conceptualized two complementary latent variables underlying patient-reported outcomes as: "value of living independently" and the "visual ability of independent living." The "value of living independently" construct can be thought of as the impact visual impairment has on the individual's life. One approach to measuring this construct is seen in the development of the Impact of Vision Impairment (IVI) scale, used to determine the effects of vision loss on adults and how they function in society.[33–35] "Visual ability" can be thought of as determining how much difficulty a patient has in achieving the visual goals defined in the Activity Inventory developed by Massof et al.,[36] "value" can be estimated by rating "importance" as the response, and "frequency" also reflects importance; "ability," on the other hand, can be estimated by rating "difficulty" as the response.[37] This difficulty rating reflects the functional reserve,[32,38] defined as the difference between the visual ability required to complete a task and the visual ability of the individual. Visual Functioning Questionnaires (VFQs) that measure "ability" rely on rating scale responses indicating the perceived level of difficulty of a task (e.g., impossible, very difficult, not difficult).[39,40] Such ordinal responses provide a hierarchy or order to the activities probed, but do not give quantitative information as to the difference between them.[41] Individuals differ from one another in terms of both the importance and the difficulty they assign to different tasks. Importance and difficulty are complementary measures underlying patient-reported outcomes and are important constructs to include in VFQ development.

PRO VFQs[42] vary by intended population, item content, length, functional requirements (visual acuity [VA] range), eye disease, and dimensions probed.[28,36,43–45] Very few VFQs have addressed the measurement of "ability" and "value" in a severe vision loss population. Finger et al.[46] developed the Impact of Vision Impairment in Very Low Vision (IVI-VLV) instrument aimed at individuals with visual acuity in the better seeing eye of less than 20/200 or visual field diameter less than 10°, or both. The IVI-VLV contains items probing the "impact" of vision loss, on ADLs and VRQoL. However, a VFQ specifically measuring "visual ability" in nearly blind individuals (i.e., how well a person with profound vision loss, including those receiving a sight-restoring therapy, can perform activities using remaining vision and without assistance) would supplement this work.

In practice, there is a need for the development of both visual function measures and self-report functional vision instruments that can quantify the vision of prosthesis wearers and those volunteering for phase 1 trials of novel treatment modalities, at baseline and follow-up, as well as those with "native" profound vision loss. The range of vision afforded by artificial and sensory substitution devices, allowing rudimentary recognition of shapes, movement, and sources of light, is not adequately captured by standard clinical definitions.[16] Therefore, the term "ultra-low vision" (ULV) was coined.[16,47] Operationally, ULV is defined as, "vision so limited that it prevents the individual from distinguishing shape and detail in ordinary viewing conditions." This definition corresponds with two categories in the International Classification of Diseases (ICD-9/ICD-10) introduced by the World Health Organization (i.e., profound visual impairment [VA $\leq$ 20/500] and near-total blindness [VA $\leq$ 20/1000]. In our research, and as explained in a companion paper (Adeyemo et al., TVST, In Press, 2017), we included individuals with reported VA less than or equal to 20/500 in the better seeing eye, but also include the restored visual ability of retinal prostheses wearers.[10] Admittedly, VA can be measured on the Early Treatment Diabetic Retinopathy Study (ETDRS) chart at 0.5 m down to 20/1600, implying that those with vision in this range can still see shapes. There is a major difference between the ideal situation of 100% contrast optotypes shown in a darkened exam room and the real world of uneven illumination and lower contrast, and one can think of a gradual changeover in vision below 20/500, where subjects may still be able to identify high contrast shapes such as optotypes, but where recognition of real-world objects, shapes, and details becomes increasingly difficult. "Counting fingers" is a case in

point: people with 20/1600 ETDRS VA cannot count fingers at 0.5 m, and even those with 20/500 may find it challenging, depending on illumination and contrast. In addition, it should be noted that the remaining visual field may be patchy, so those with severe vision loss may be using islands of vision, contributing to their difficulty seeing shape and details.

VFQs measuring "ability" rely on rating scale responses indicating the perceived level of difficulty of a task (e.g., impossible, very difficult, not difficult).[40,39] Such ordinal responses provide a hierarchy or order to the activities probed, but do not give quantitative information as to the difference between them.[41] Modern day psychometric techniques such as Rasch analysis convert these ordinal response data to an interval scale that meets customary criteria for measurement.[41,48,49] Interval scales allow both order and measurable increments between responses (e.g., time, temperature).[41] This effectively calibrates the scale formed by the items and the respondents, and allows meaningful comparisons across individuals, time, and treatment interventions. In other words, a calibrated VFQ quantitatively ranks persons according to their level of ability, and items according to level of difficulty. Rasch analysis also takes into account the inherent diversity among participants. Due to an individual's differences in remaining vision, preferences, and experiences, not all items will be applicable, and one can allow respondents to skip those items. Rasch analysis accommodates missing data and reports how well the individual items and response sets match the theoretical model, how much of the total variance in the data is explained by the model, and the dimensionality of the latent trait measured by the instrument (i.e., whether the targeted construct, such as visual ability, is itself made up of more than one independent variable).[41]

The mathematical approach contained in Rasch analysis critically depends on the item difficulties and person visual abilities spanning matching ranges, so the scores assigned by the respondents cover the full range and vary among individuals. For this reason, a VFQ for use in ULV has to be calibrated in a ULV population. A different way of expressing this is the requirement that the instrument have content validity, that is, the items address activities that are informative about visual ability, and measurement validity (i.e., the items, persons, and responses form a coherent data set under the mathematical model). To develop a new VFQ specifically for use in a ULV population, we therefore need items describing visual

activities that are representative of those carried out by individuals with ULV, and we have collected a large set of such activities. In a separate publication, we report the results of a focus group study, conducted as the first stage of the Prosthetic Low Vision Rehabilitation (PLoVR) program development effort (Adeyemo et al., TVST, In Press, 2017), and aimed at creating a broad inventory of visual activities characteristic of ULV (i.e., activities that require very limited vision), but cannot be accomplished blindly. The 760 visual activities reported by the focus group members in that study could be categorized into functional visual domains (reading/shape recognition, mobility, visual-motor, and visual information), but were found to be more appropriately characterized by visual aspects such as contrast, lighting, motion perception, familiarity, size, and distance. ULV participants in the focus groups described their ability to accomplish a task as dependent on these visual aspects. For example, whether a task is achievable with the person's rudimentary vision may depend on figure-ground contrast, such as black coffee in a white coffee mug. Thus, creating content valid for the ULV population requires taking these visual aspects into account, and a representative distribution of visual aspects as well as visual domains was a guiding principle in the choice of items for a VFQ targeted at individuals with ULV. In terms of visual domains, given the limitations of ULV the "reading" domain is reduced to "crude shape recognition" in this population, and will be designated as "shape" in the remainder of this paper.

Here, we report on the second stage in the PLoVR program development (i.e., the creation and psychometric evaluation of a VFQ in a ULV population) called the ULV-VFQ. The advent of retinal prostheses and other sight-restoring therapies has made the development of such an instrument particularly timely, but the ULV-VFQ also represents the first PRO instrument specifically targeting "ability" in individuals with native ULV as low as bare light perception, allowing estimation of their visual ability and offering the opportunity to measure their visual ability as their vision evolves and potentially benefits from training or novel treatments.

Methods

This study was conducted between January 2013 and March 2014. Here, we describe the generation of items used in the questionnaire, the administration of

the items to ULV participants and the psychometric evaluation in an iterative process, as steps in the development of a calibrated instrument. The study protocol was approved by the institutional review board of the Johns Hopkins University School of Medicine and followed the tenets of the Declaration of Helsinki. All participants gave informed consent for study participation after explanation of the nature, possible risks and benefits of the study.

Participants

Participants were recruited from the Low Vision Clinic of the Wilmer Eye Institute at Johns Hopkins Hospital (Baltimore, MD), through advocacy groups (e.g., Foundation Fighting Blindness, the National Federation of the Blind) and online forums for the blind and visually impaired, and through email and flyers to eye care providers and prospective participants nationwide. Participants had to be at least 10-years old, be proficient in spoken English, and meet our only visual inclusion criterion: a history of vision in the counting fingers/hand motions/light perception range (VA $\leq$ 20/500 in the better eye) as determined by medical records.

In Round 1 of the survey administration, there were 90 participants. All participants were legally blind with VA less than or equal to 20/500; most had lost form vision, with 73% reporting usable vision at the time of survey administration. Our sample was comprised of 67% participants with inherited retinal degenerations, including six retinal implant recipients who took part in the Argus II feasibility study.[50] The same participants were included in Round 2, with the exception of 10 participants who opted not to participate or were not available during the second round of administration. Considerable effort was required to recruit qualifying participants, as ULV individuals represent a small subset within the legally blind population[51]; many prospective participants were screened out (e.g., if they indicated being able to read magnified print).

Procedures and Instrumentation

Questionnaire items were developed and written by two authors (PEJ and CR) based on the functional activities described in the focus groups, and reviewed and refined by all members of the study team. Items were tagged by category according to functional visual domain and visual aspect (Adeyemo et al., TVST, In Press, 2017) to ensure a representative distribution of items across all categories. Generally, items were worded as: "How difficult is it to [do this activity] using your remaining vision?" We emphasized using only vision, without or with the use of visual assistive devices, such as a magnifier (see Supplementary Materials S1 for instructions and items).[52] The ordered difficulty ratings were similar to those used by Massof et al.[36] during the development of the activity inventory and consisted of: not difficult, somewhat difficult, very difficult, impossible to perform visually, and not applicable. Items were uploaded to Survey-Monkey (SurveyMonkey, Inc., Palo Alto, CA, USA), which was then used by a research assistant for administration over the phone, or by the participant for online self-administration, according to preference. Participants with access to computers who were comfortable with the use of voiceover and other speech output software typically elected to respond online. According to the SurveyMonkey log, questionnaire administration took less than 1 hour for both phone and self-administered routes. Two rounds of the questionnaire were administered, with a revision after Round 1 based on psychometric properties and participant feedback. Additional information was collected each round to gauge current usable vision as reported by each participant.

Psychometric Analysis

Response data were recoded numerically in inverse order of difficulty level, with "not applicable" coded as zero so as not to be included in the analysis (note: the number of "not applicable" responses was small [4.2% in Round 2]). Rasch analysis was performed using Winsteps (v. 3.92.1; in the public domain, www. winsteps.com), applying the Andrich polytomous model for joint maximum-likelihood estimation[53] in order to convert ordinal data (i.e., recoded difficulty rating categories) into an interval scale.[54]

Person-item maps were plotted to represent item difficulty and person ability measure estimates along a common interval scale in logits (i.e., log odds ratio unit),[54–56] with mean item difficulty at 0 (by convention).[54] This person-item map was used to compare the distributions of item difficulty and person visual ability estimates.[55] In order to test content and measurement validity of the estimated measures, we used fit statistics, allowing us to examine how well the Rasch model predicts the response of each respondent to each item.[55] Mean square fit statistics were calculated to evaluate the assumption of unidimensionality. Infit (Information-weighted sum) and outfit (outlier sensitive mean squares) quantities were computed to determine reliability of the latent variable measure-

ment.[49,54] Person reliability was calculated to evaluate the replicability of person ordering that we would expect if the same sample population were given a parallel set of items. Item reliability was calculated to determine item placement consistency along the variable axis, should these same items be given to another sample population of similar characteristics.[54] Principal component analysis (PCA) was conducted on the second round residuals to determine if the latent trait is one-dimensional (i.e., if more than one latent variable would be required to explain the variability).[57] Differential item functioning (DIF) tests were carried out to examine whether certain categories of respondents (e.g., Argus II users; those responding online; those who report that they no longer use their vision) judge item difficulties differently from the remaining respondents, which would refute the hypothesis of a common visual ability variable. Similarly, differential person functioning (DPF) was applied to categories of items that might assign different ability rankings to some categories of respondents.

Results

Participant Demographics

Table 1 describes the demographic, socioeconomic, and other characteristics of the total participant sample ($n = 90$). The median age was 61 years (range, 24–103 years), and 47% of the participants were female. Remarkably, even though this was in no way a selection criterion, 78% of all participants had a form of inherited of congenital retinal dystrophy.

Content Area and Item Identification

Table 2 shows the distribution of items according to content; this distribution follows that identified in the focus group responses (Adeyemo et al., TVST, In Press, 2017). Note the large number of items characterized by high contrast (~60%) and lighting (~14%). This method of item generation yielded 149 questions for the first round questionnaire and 150 for the second round questionnaire. Items were worded at an eighth grade reading level according to a Flesch-Kincaid Grade Level Test.

Round 1–Survey Analysis

Person-Item Mapping

In Figure 1, a histogram of estimated person logit scores is plotted to the left side of the vertical axis,

with a discernable person measure range from −6.0 to +2.8 logits; the item score histogram is plotted to the right side, with an item measure range from −3.3 to +3.0 logits. The ordering of estimates and the assignment of relative values via logit scaling are essential features of measurement.[41] More difficult items and more able persons are located toward the top (i.e., >0), while the less difficult items and less able persons are located toward the bottom (<0) of the scale. The mean of the person distribution is offset by −1.55 logits relative to the mean of the item responses (0.0 logits, by convention). Relative to the 9-logit range of person measures, the items in this sample showed a range less than 6.5 logits and were moderately well-targeted, according to criteria formulated by Pesudovs.[56] From this figure, it is evident that the initial version of the instrument was able to discriminate between less and more able participants. However, when comparing the distributions of item and person scores, there were only approximately five items in the "less difficult" range to map the approximately 25 "less able" persons with logit scores less than −2.50, suggesting that additional easy questions were needed, either by substituting easy questions for more difficult ones, or by modifying some items to make the described tasks easier.

Fit Statistics

The reliability index was 0.97 for items and 0.99 for persons in the Round 1 analysis, indicating that the model fit is of sufficient quality for the estimates to be interpreted as interval level measures. A mean value of 1 was found for the infit and outfit mean squares, indicating that the items and persons contributed reliably to the measurement of the latent variable (i.e., visual ability) and showing that the data are compatible with the Rasch model.

Standardized fit statistics (i.e., z-scores) of estimated person measures in Round 1 are shown in Figure 2 (Person Bubble Chart [top]) with each bubble diameter corresponding to the standard error of the person measure estimate. In the first round ($N = 90$), person estimates were compatible with the model with the exception of three participants (outliers) with z-scores greater than 4 (underfit), suggesting that their responses are less predictable than expected by the model. Upon inspection of the individual data, nothing of note was determined for these individuals' degree of vision loss, diagnosis, or age. Overfitted person estimates do not add any new information to the model fit: they are more predictable, but this is of no concern.

Item measures for Round 1 are plotted in Figure 2

Table 1. Characteristics of Sample as Mean ± SD or *n* (%)

Total Sample, Round 1 (*n* = 90)	Value
Age	61 ± 15
Sex	
Female	42 (47)
Male	48 (53)
Group	
Phone	63 (70)
Online	27 (30)
Usable Vision	
Yes	66 (73)
No	24 (29)
Visual Impairment	
<20/400 to >CF	5 (6)
CF to >LP	24 (27)
LP and worse	55 (60)
Retinal implants (LP)	6 (7)
Diagnosis	
Retinitis pigmentosa	60 (67)
Age-related macular degeneration	3 (3)
Other retinal dystrophy	10 (11)
Glaucoma	7 (8)
Other	10 (11)
Education	
High school	8 (9)
Some college or technical school	28 (31)
Completed college or technical school	17 (19)
Some graduate/professional school	7 (8)
Completed graduate/professional school	26 (29)
Did not respond	4 (3)
Race	
Caucasian	76 (84)
African-American	5 (6)
Hispanic/Latino	3 (3)
Asian	3 (3)
Other	2 (2)
Did not respond	1 (1)
Annual Income	
≤20,000	18 (20)
$21,000–$40,000	16 (18)
$41,000–$60,000	14 (16)
$61,000–$80,000	11 (12)
>$80,000	21 (23)
Did not respond	10 (11)
Employment	
Employed	23 (26)
Self-employed	12 (13)
Retired	35 (39)

Table 1. Continued

Total Sample, Round 1 (*n* = 90)	Value
Semi-retired	1 (1)
Disabled	1 (1)
Unemployed	6 (7)
Volunteer	1 (1)
Student	4 (4)
Did not respond	7 (8)
Relationship	
Single, never married	15 (17)
Living with partner, not married	3 (3)
Married	42 (47)
Divorced or separated	14 (16)
Widowed	6 (7)
Did not respond	10 (11)
Living Situation	
Alone	28 (31)
With family or friends	58 (64)
Did not respond	4 (4)
Activity	
Active	71 (79)
Sedentary	14 (16)
No or ambiguous response	5 (6)

CF, Counting Fingers; LP, Light perception. This table describes the characteristics for Round 1 of the questionnaire development (*n* = 90).

(item bubble chart; bottom). Four items (outliers) were estimated beyond the z-score cut-off of 4 (underfit). Upon inspection, it was determined that the two closer outliers represented ambiguous items (numbers 2 and 76; see Supplementary Materials S1) and could be reworded. The two farthest outliers queried the respondent's ability to adjust to changes in lighting (e.g., entering into a dark building on a sunny day), items that were inherently variable, because they are known to be more difficult for retinitis pigmentosa patients than for others. This may be important information, and because only two items were affected it was decided to retain these two items for Round 2 administration of the questionnaire.

DIF and DPF Analysis

To determine if there were any differences in responses between subsets of participants, we performed a DIF analysis along three dimensions: (1) online versus phone administration, (2) remaining usable vision versus self-reported functional blindness, and (3) Argus II use versus native ULV. Because the DIF analysis was

Table 2. Content Area and Item Identification

| Item Codes | Visual Aspect | Visual Aspects | | | |
| | | Round 1 | | Round 2 | |
		# of Items	% of Items	# of Items	% of Items
CO	Contrast	88	59	92	62
LM	Luminance	16	11	19	13
EL	Environmental lighting	5	3	3	2
FM	Familiarity	8	5	5	3
MP	Motion perception	9	6	8	5
DI	Distance	9	6	10	7
SZ	Size	4	3	3	2
MI	Other/miscellaneous	10	7	10	7

| Item Codes | Functional Dimension | Functional Dimensions | | | |
| | | Round 1 | | Round 2 | |
		# of Items	% of Items	# of Items	% of Items
RE	Reading	3	2	3	2
MO	Mobility	14	9	10	7
VM	Visual motor	31	21	30	20
IG	Information gathering	101	68	107	72

* Data based on focus group study published in Adeyemo et al. (TVST, In Press, 2017) and used to identify content for item generation.

carried out on 149 items, a Bonferroni correction was applied, changing the required *P* value from 0.05 to 0.0003; therefore, even though on each of the three analyses several items showed raw probabilities below *P* equal to 0.01, no items reached significance, and no evidence for model differences was found along any of the three dimensions examined.

In a similar analysis, performed to examine the two misfitting items dealing with light and dark adaptation for DPF, no significant difference was found.

Item Revisions for Round 2

Based on participant feedback in the comment sections during the first round, some items were flagged as ambiguous. For example, in Round 1 of the survey, one question (number 48, see Supplementary Materials S1) was, "When playing a piano, how difficult is it to see the difference between the black and white keys?" However, several participants noted that this depends on the lighting in the room. Because we wanted this to be a high contrast question, we revised the sentence to rule out any ambiguity due to lighting. The revised question read, "When playing a piano in a well-lit room, how difficult is it to see the difference between the black and white keys?" In most cases, the

visual aspect categorization stayed the same (Table 2 and Supplementary Material S1), but the item needed precision in order to hone in on the visual aspect of interest. In general, items were flagged for revision if they were considered imprecise, redundant, or unnecessarily visually demanding as judged by consensus between two authors (PEJ and CR). In some cases, demanding items were removed and simpler ones were inserted based on other activities reported by the focus groups (Adeyemo et al., TVST, In Press, 2017). With these considerations in mind, adjustments were made to a large majority of the items; most of them minor, and a second version of the questionnaire with 150 items was generated and administrated to the participants (see Supplementary Materials S1 for all items).

Round 2 Survey

Person-Item Mapping

In Figure 3, estimated person scores (on the left) ranged from −5.1 (less able) to +4.4 (more able) logits, suggesting that the instrument was able to differentiate reliably among different abilities, with a distribution mean of −0.86. The item distribution spanned

Map of Persons and Items
Round 1

More Able .... **More Difficult**
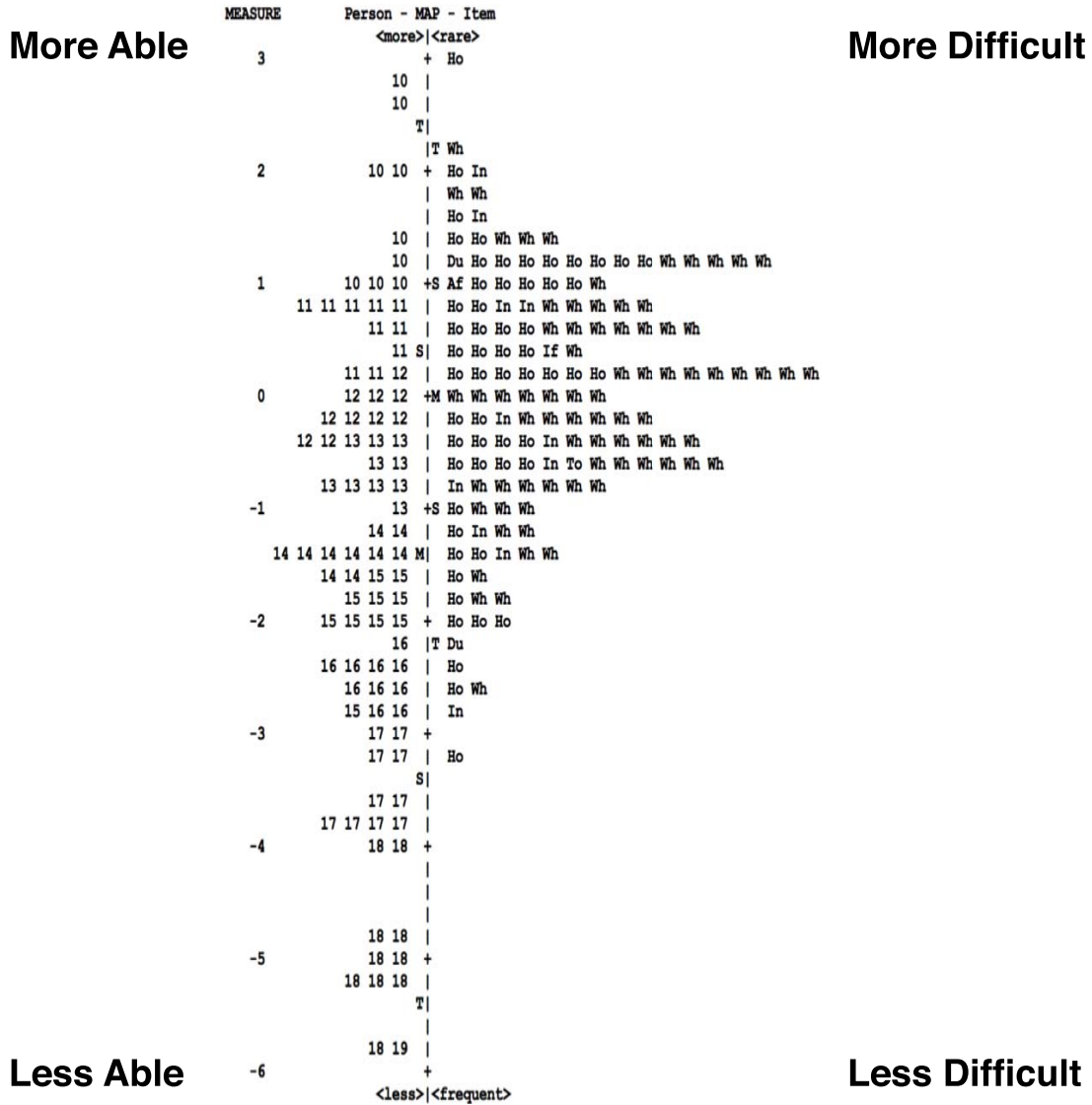
Less Able .... **Less Difficult**

Figure 1. Person-Item Map (Round 1).

a range from −2.7 (less difficult) to +3.4 (more difficult). The item distribution is well matched to the person distribution, within less than 1 logit difference between the means.[56]

Fit Statistics

The reliability index in Round 2 was 0.97 for Items and 0.99 for Persons, similar to those in Round 1. In standardized test statistics for Round 2 person measures (Fig. 4; top), six of the participants had estimated response z-scores greater than 4, suggesting that their responses were less predictable (i.e., more variable). In standardized statistics for the item measures (Fig. 4; bottom), four items were estimated beyond a z-score of 4 including the two items querying the ability to adjust in sudden light changing situations. Overall, the second round of questionnaire showed good performance. In all cases, the effect of the outliers on the overall fit was negligible as determined by the PCA described below.
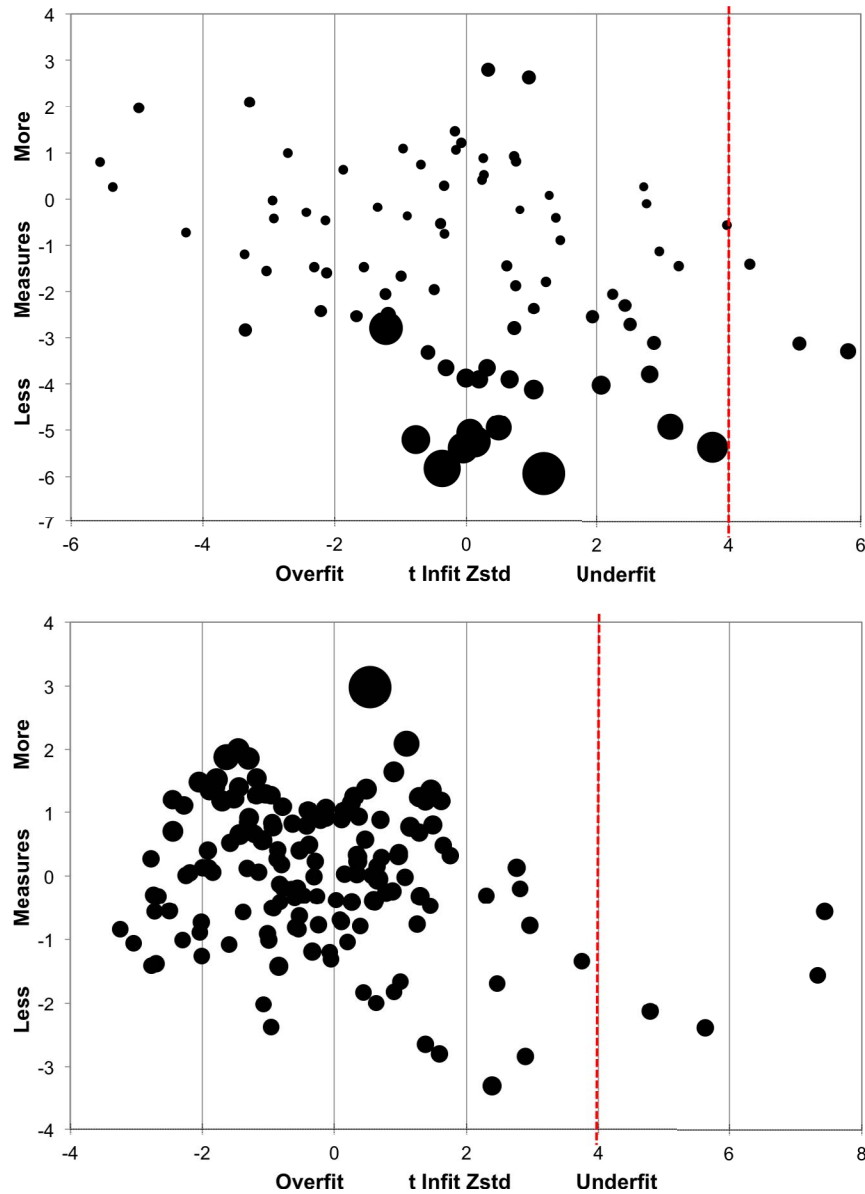
Figure 2. Standardized fit statistics (Round 1): person (*N* = 90; *top*) and item (*N* = 149; *bottom*) measures.

Visual Domains and Aspects

In Figure 5, Round 2 item measures are replotted with color codes according to visual aspects (top) and visual domains (bottom). The visual aspects distribution supports participants' descriptions of what allows them to accomplish a particular task: items determined by luminance and illumination are easier (lower in the figure), while those determined by familiarity, movement, size, and distance require more vision (higher item measures); items determined by contrast span a wider range. In the visual domains distribution, the "shape" items are confirmed to be more difficult than those falling in other domains.

Principal Component Analysis (PCA)

The explained variance in Round 2 was estimated at 69.9%, and PCA was conducted on the residuals (i.e., the unexplained variance). This 30.1% of the variance represents measurement noise; this noise may be random, but it may include systematic deviations from the model, and PCA can uncover such deviations. As Figure 6 shows, the first principal component of the residual variance explains only 2.6% of the total variance, and each component after that explains less. The absence of major systematic deviations from the Rasch model indicates that the data set is internally consistent, with a single latent

Map of Persons and Items
Round 2

```
TABLE 1.0 rasch.plovr.3.10.14                ZOU504WS.TXT  Mar 11 22:35 2014
INPUT: 81 Person  150 Item  REPORTED: 80 Person  150 Item  4 CATS WINSTEPS 3.80.1
--------------------------------------------------------------------------------

        MEASURE                  Person - MAP - Item
                                  <more>|<rare>
More Able   5                         +                          More Difficult
                                      |
                                      |
                                  126 |
                                      |
            4                         +
                                      |
                                  113 |
                                  127 T|
                                      |
            3                         +
                                      |
                                      |   7 7
                      134 146         |T
                      129 137 160     |
            2                     143 +   5
                                  166 |   0 1 5 5 6 6 8
                                      |   1 1 3 3
                                      |   3 3
                      120 132 S|S 0 1 2 4 5 7
            1                     121 +   0 1 1 1 2 3 3 5 5 6 7 7 8 8
                  112 123 180         |   2 3 4 5 6 7 7 7 7 8
           109 140 149 151 172        |   0 1 2 4 7 8
                  102 135 147         |   0 1 1 3 3 4 4 5 6 6 7 7
                                      |   0 1 1 2 4 4 5 6 7 8 8
            0             101 110 150 159 +M 0 0 0 0 1 2 2 2 4 4 4 5 5 6 7 7 8 8
       105 117 125 145 148 153 154 177   |   3 4 5 6 8 8
                                  152 |   2 3 4 6 6
                      104 119 158     |   0 1 2 2 2 4 5 6 6 8
                      118 144 M|   0 2 3 3 4 9
           -1   124 128 130 156 161 173 178 +   0 1 2 3 7 9
                                  169 |S 0 2 8
                      111 138         |   0 3 5
                                  133 |   4 5 6 6 8
                      115 170         |   1 5
           -2         103 122 174     +   8
                      106 141 142     |   4 8
                                  155 |T
                                  176 |   2 3 6 9
                                  165 |   9 9
           -3         116 164 167 S+   9
                                  136 |
                                      |   9
                                  175 |
                      107 108         |
           -4         171 179         +
                                  181 |
                                      |
                      157 168         |
                      114 162 163     |
           -5                         +
                                  131 T|
                                      |
Less Able   -6                        |                          Less Difficult
                                      +
                                  <less>|<frequent>
```
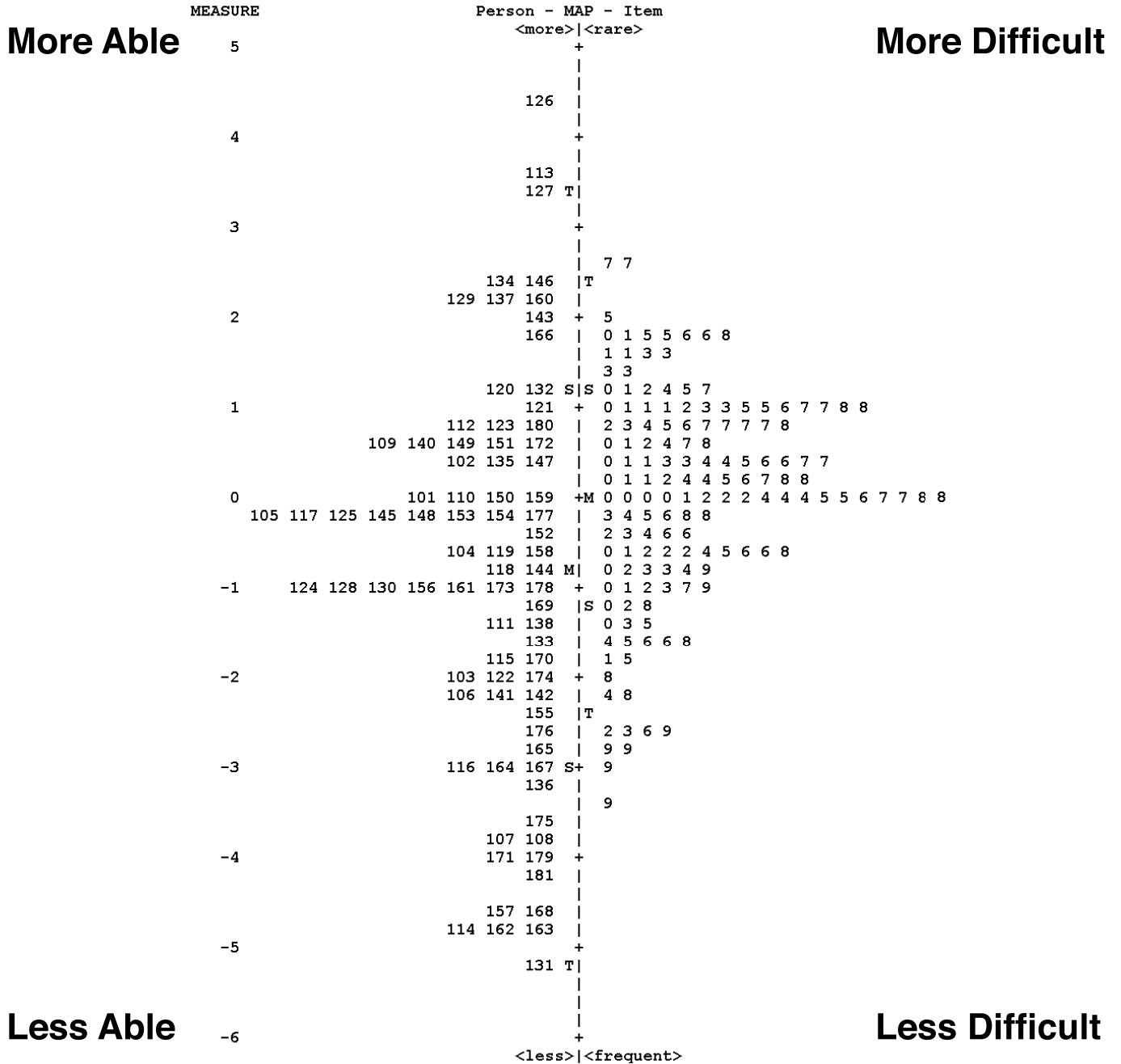
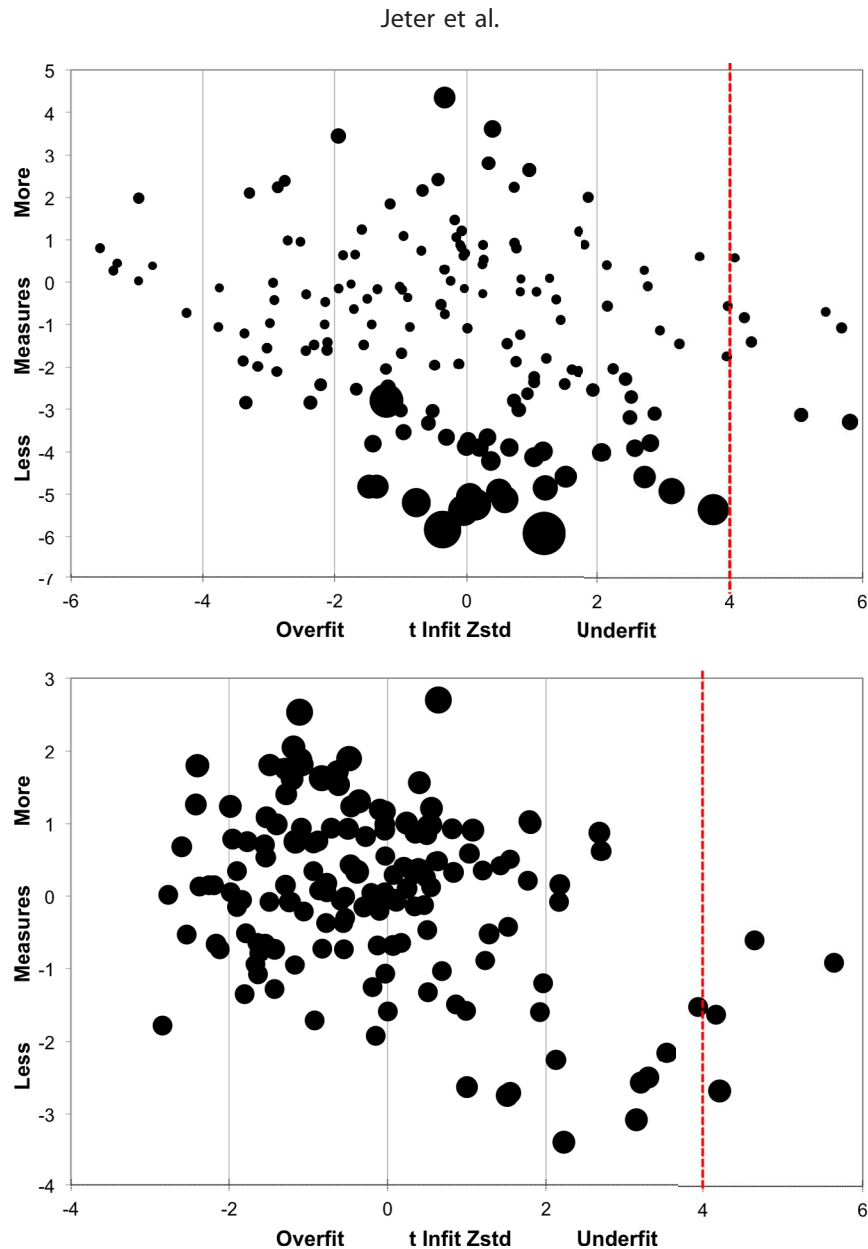translational vision science & technology

Figure 4.   Standardized fit statistics (Round 2): person ($N = 80$; *top*) and item ($N = 150$; *bottom*) measures.

variable, and that the instrument yields a valid measurement.

Discussion

Using content elicited from a ULV population (Adeyemo et al., TVST, In Press, 2017), including six Argus II Retinal Implant wearers, we developed a self-report questionnaire, the ULV-VFQ, as part of the PLoVR program development. Items were classified according to visual aspects (e.g., contrast) required to complete the task and according to visual domains, and reflect activities that individuals with

ULV are able to accomplish with their remaining vision including visual assistive devices, but which they could not accomplish based solely on other sensory input. The instrument characteristics were psychometrically evaluated using Rasch analysis and found to meet the criteria for measurement that include hierarchical ordering and interval scaling[41]; moreover, the latent variable underlying the responses was determined to be unidimensional, and can be characterized as visual ability. The ULV-VFQ covers a wide range of item difficulties and can differentiate among different levels of ability among individuals with ULV. Fit statistics were determined to be
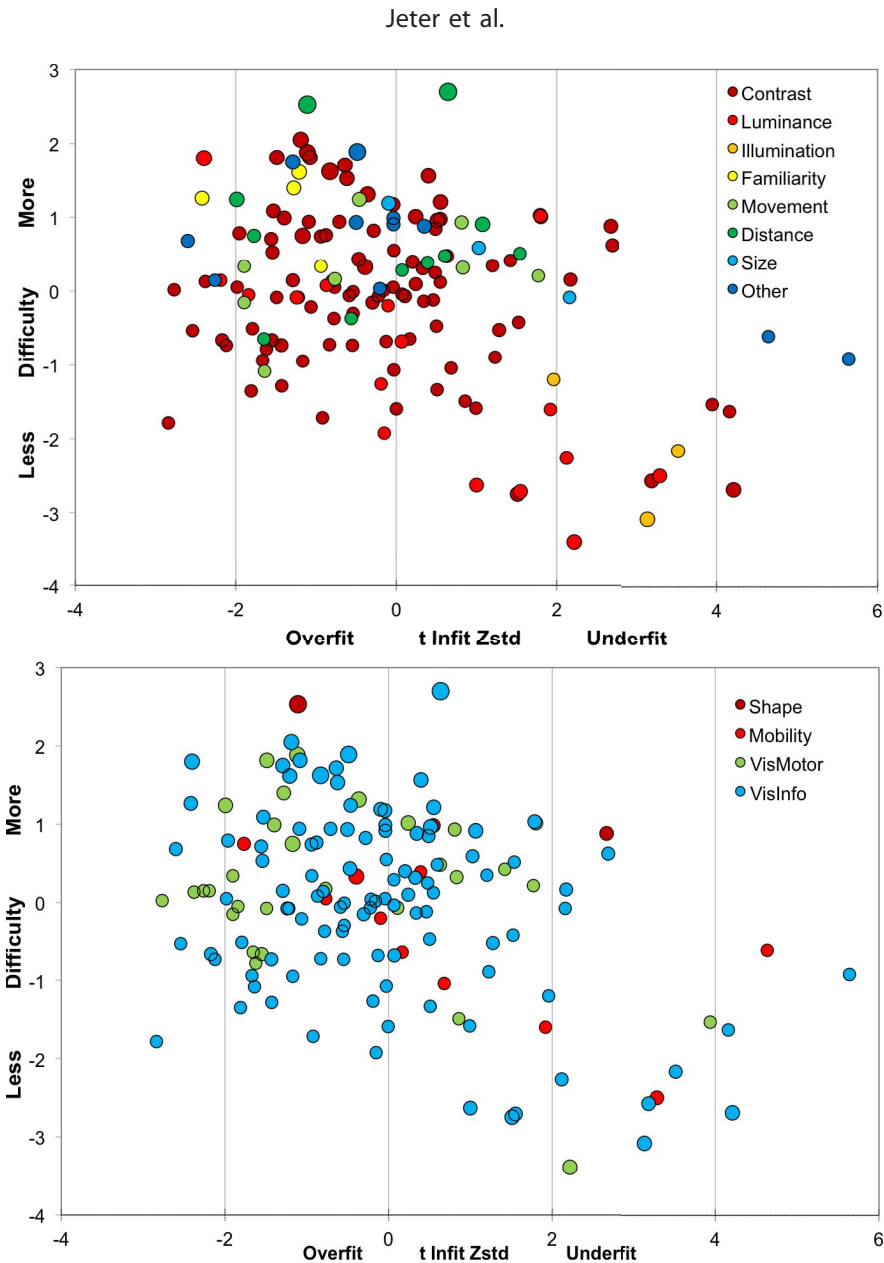
Figure 5.   Round 2 item measures as visual aspects ($N = 150$; *top*) and domains ($N = 150$; *bottom*).

excellent, and only a few items were underfit; PCA showed that the hypothesis of a unidimensional latent variable is sustained by the responses from our sample of 80 to 90 ULV individuals.

Because of the generally good fit statistics and the absence of DPF, we decided not to eliminate any underfitted items, including the two items dealing with light and dark adaptation. This decision was taken on advice of coauthor RWM, with the argument that retaining these items would have a negligible effect on the overall fit in the Rasch analysis and that, even if they might show DPF for RP patients versus others in a larger population of

individuals with native and restored ULV, having these items may be a useful measure to classify individuals with ULV. In general, there is a downside to eliminating misfitting items, especially if they are felt to probe important information.

For similar reasons, and supported by the absence of DIF for any subgroups in our subject population, we feel that we are justified in basing the best estimates for the 150 item measures in the ULV-VFQ on two modes of administration (phone and online), a mixture of currently and formerly sighted individuals, and a subset of individuals with an Argus II retinal prosthesis.
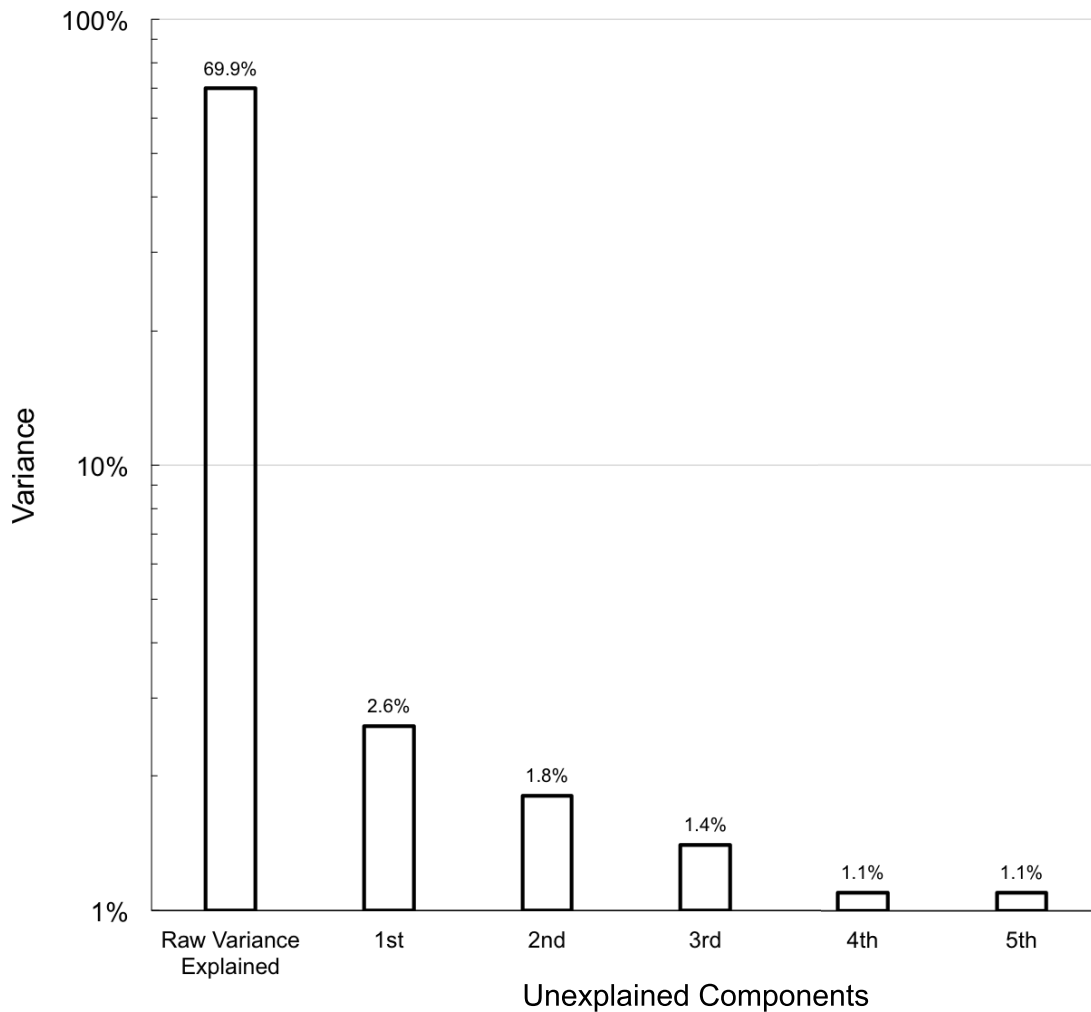
Figure 6. Principal component analysis.

There are several limitations to the study that are worth of mention. Despite casting a wide net to recruit as many participants as possible, we found that ULV individuals were difficult to access. This is supported by the low prevalence of ULV among the visually impaired served by low vision clinics reported in a multicenter study[51]: among 646 participants, only 39 were within the ULV range. In this study, we were able to recruit 90 eligible participants. The majority of our participants were diagnosed with an inherited retinal degeneration, but other diseases were also represented. We have no way of knowing whether the high percentage of retinal degeneration in our sample is representative of the ULV population in general: none of the blindness statistics provided by national or state registries differentiate between legal blindness with form vision, legal blindness without form vision (i.e., ULV), and total blindness. Most of our participants had end stage vision loss, and we thus

had no way of knowing if they had predominantly peripheral or central areas of remaining vision, unless it was self-reported. Additionally, we were unable to determine of the presence or absence of color vision in our population. However, participants in our focus groups study (Adeyemo et al., TVST, In Press, 2017) all reported that the appearance of color has been reduced to brightness differences, in almost all situations. Our focus in choosing items for the ULV-VFQ was on what our population was able to do with their remaining vision. Regardless of the actual distribution of the causes of vision loss, both our focus group study (Adeyemo et al., TVST, In Press, 2017) and the current study show that individuals with ULV do use their remaining vision, but depend on high contrast and good lighting for it to be effective. Because our study population was recruited from locations across the country (USA), we relied on eye examination records to estimate visual

function. Quantifying visual measures at the level of ULV is difficult, in any case. We did not collect specific information about comorbid conditions, other than verifying that the participants were in good general health and had no trouble understanding the instructions and questions.

Questionnaire items were worded at an eighth grade reading level according to a Flesch-Kincaid Grade Level Test analysis; however, it is possible that variation in item readability existed within the instrument.[58] More importantly, the directions for the questionnaire were long and written at the 11th grade level. Most of our participants (∼86%) reported a greater than high school education, and the wording in those instructions would be familiar to an educated person with longstanding visual impairment. Furthermore, most of our participants responded by phone (70%) and the test administrator was able to clarify any confusing language in the instructions. Finally, the consistency in responses suggests that participants understood the instructions; any misunderstandings would have introduced additional variance into the data.

Despite our hope to recruit additional participants in visual prosthesis and gene therapy trials, the only such participants in our sample were Argus II users available at the time of Round 1 administration (i.e., prior to Argus II approval by the FDA). We did not have access to a population participating in other vision restoration treatments. Argus II users in our sample showed no significant difference in their ULV-VFQ responses from other participants according to a DIF analysis; in other words, our data provide no evidence that artificial sight is governed by a different kind of visual ability than native rudimentary vision. Further studies that include a larger cohort of retinal implant wearers and patients regaining vision through other treatment modalities will be required to explore potential differences between native and restored ULV. Incidentally, our data also indicated that neither the route of administration (online versus phone) nor the presence versus absence of residual usable vision affected the latent visual ability trait.

Commonly used VFQs[36,44,45,59–61] vary by targeted visual disorder, age group, and so on, and are primarily aimed at the range of low vision that includes intact shape perception (i.e., better vision than ULV), with the notable exception of the IVI-VLV,[46] aimed at individuals with VA less than 20/200, and responsive well below that. As reported earlier, the IVI-VLV queries respondents how much vision loss has "interfered" with their ADLs and VRQoL (e.g., "How much does your eyesight interfere with crossing the street?")

reflecting a difference in the underlying latent variable measured, "impact" versus "ability." Because the two instruments measure a different latent variable, they are not equivalent, even if the subject populations were similar. The difference between the ULV-VFQ and IVI-VLV is perhaps best illustrated by comparing the least demanding item in the IVI-VLV "Activities of Daily Living, Mobility & Safety" subscale (i.e., "How much does your eyesight affect your general safety getting around at home?") to the corresponding item "When standing outside at high noon, how difficult is it for you to tell whether it is sunny outside?": while the IVI-VLV item probes limitations experienced getting around in a familiar environment (which may not require any sight), the ULV-VFQ specifically asks about the ability to see. This difference in item focus between the IVI-VLV and the ULV-VFQ causes the two instruments to be complementary rather than redundant. Both instruments have excellent psychometric properties: the person and item reliability indices are all greater than 0.90. A future comparison in an appropriately chosen sample will be very interesting and indeed, a holistic approach to vision rehabilitation is preferred. Additional comparisons with the IVI-LVL[46] are worth mentioning. Our preference was to measure a tightly defined construct ("visual ability"), whereas the IVI-VLV measures two separate constructs: Activities of Daily Living, Mobility and Safety (ADLMS) and Emotional Well-being (EWB). We estimate ability by measuring difficulty (as a response category), whereas the IVI-VLV[46] uses other response categories such as "frequency" (e.g., "a lot"), which are potentially contaminated because it is not known how often any given task needs to be completed by a particular individual to be qualified as "a lot." Finally, the IVI-LVL instructs participants to "Please answer about YOUR eyesight or activities with GLASSES, any AIDS and/or DEVICES, if you use them." These instructions do not stipulate that these assistive devices must be visual, and thus potentially introduce another confound, whereby a participant completes the task using other sensory input, such as tactile information when walking with a cane. In this example, it would be difficult, if not impossible, to discern what a participant is able to achieve with vision alone, which was the primary goal of the ULV-VFQ.

The ULV-VFQ probes how much participants are able to accomplish with their remaining vision. An important goal of the PLoVR development is to design instruments that track disease progression and changes in vision after visual restoration treatment and rehabilitation. We specifically designed the ULV-

VFQ to measure functional reserve[38] by querying respondents' ability to perform tasks visually, under precisely defined conditions. Thus the question: "How difficult is it for you to see a freshly-painted white crosswalk on dark pavement on a cloudy day?" seeks to eliminate ambiguity about the conditions and implies contrast as the specific visual aspect that was determined through earlier focus groups to be critical for successful task performance. In this example, if a patient received a retinal implant, end-stage vision loss is an eligibility criterion. The patient with end-stage vision loss is likely to already have orientation and mobility (O & M) skills. Here, we are most interested in the remaining vision available pre-implant versus that available post-implant and to measure that, we need a specific tool targeting remaining vision only, as nonvisual O & M skills are not likely to change dramatically. Whether O & M skills will improve following vision restoration will depend on the change in visual ability, so it is critically important that we are able to assess this separately from other abilities. This does not diminish the need to take non-visual abilities into account in a comprehensive rehabilitation approach. In that sense, the IVI-LVL and ULV-VFQ can complement each other in providing different perspectives on the patient's progress. What is important is that each instrument is cognizant of the latent variable it measures.

The ULV-VFQ was administered at a single timepoint (i.e., cross-sectionally), and additional studies will be needed to determine the responsiveness of the instrument to treatment effects or disease progression. In its present form, it is also not ready for use in a clinical trial or rehabilitation program: a questionnaire with 150 items is lengthy, and items with similar difficulty may be redundant. Shorter[62] and adaptive[63] versions of the ULV-VFQ have been developed and are better suited for this purpose. On the other hand, we feel that there is merit to having a large set of items at this stage of development: not only can we continue to monitor these items for optimal fit properties as data are collected in larger groups of individuals with ULV, but also will we be able to draw from questions with similar item measures when for the computerized adaptive version of the ULV-VFQ and for standardized visual performance assessment through calibrated activities.

Conclusion

In conclusion, an instrument with 150 items, called the ULV-VFQ, was developed as part of the PLoVR program development program. It was determined, by Rasch analysis, to have excellent psychometric properties, measuring a unidimensional latent trait (i.e., visual ability). Future directions include administration to a larger ULV sample, including more retinal implant wearers and beneficiaries of other vision restoration treatments, to establish validity in a diverse population.

References

1. Al-Shamekh S, Goldberg JL. Retinal repair with induced pluripotent stem cells. *Transl Res*. 2014; 163:377–386.
2. MacLaren RE, Pearson RA. Stem cell therapy and the retina. *Eye*. 2007;21:1352–1359.
3. McClements ME, MacLaren RE. Gene therapy for retinal disease. *Transl Res*. 2013;161:241–254.
4. Petrs-Silva H, Linden R. Advances in gene therapy technologies to treat retinitis pigmentosa. *Clin Ophthalmol*. 2013;8:127–136.
5. Simonelli F, Maguire AM, Testa F, et al. Gene therapy for Leber's congenital amaurosis is safe and effective through 1.5 years after vector administration. *Mol Ther*. 2009;18:643–650.
6. Sahni JN, Angi M, Irigoyen C, Semeraro F, Romano MR, Parmeggiani F. Therapeutic challenges to retinitis pigmentosa: from neuroprotection to gene therapy. *Curr Genomics*. 2011;12: 276–284.
7. Humayun MS, Weiland JD, Chader G, Greenbaum E. Artificial Sight: Basic Research, Bio-

medical Engineering, and Clinical Advances. New York: Springer; 2007.

8. Dagnelie G. Visual prosthetics 2006: assessment and expectations. *Expert Rev Med Devices*. 2006; 3:315–325.

9. Dagnelie G. Retinal implants: emergence of a multidisciplinary field. *Curr Opin Neurol*. 2012; 25:67–75.

10. Stronks HC, Dagnelie G. The functional performance of the Argus II retinal prosthesis. *Expert Rev Med Devices*. 2014;11:23–30.

11. Chader GJ, Weiland J, Humayun MS. Artificial vision: needs, functioning, and testing of a retinal electronic prosthesis. *Prog Brain Res*. 2009;175: 317–332.

12. Geruschat DR, Dagnelie G. Restoration of vision following long-termblindness: considerations or providing rehabilitation. *J Vis Impair Blind*. 2016; 110:1–5.

13. Ho AC, Humayun MS, Dorn JD, et al. Long-term results from an epiretinal prosthesis to restore sight to the blind. *Ophthalmology*. 2015; 122:1547–1554.

14. Lepri BP. Is acuity enough? Other considerations in clinical investigations of visual prostheses. *J Neural Eng*. 2009;6:1–4.

15. Wilke R, Bach M, Wilhelm B, Durst W, Trauzettel-Klosinski S, Zrenner E. Testing visual functions in patients with visual prostheses. In: Humayun MS, Weiland JD, Chader G, Greenbaum E, eds. *Artificial Sight*. Springer; 2007:91–110.

16. Nau A, Bach M, Fisher C. Clinical tests of ultra-low vision used to evaluate rudimentary visual perceptions enabled by the BrainPort vision device. *Transl Vis Sci Technol*. 2013;2(3):1–12.

17. Colenbrander A. Visual functions and functional vision. *Int Congr Ser*. 2005;1282:482–486.

18. Haymes SA, Johnston AW, Heyes AD. The development of the Melbourne low-vision ADL index: a measure of vision disability. *Invest Ophthalmol Vis Sci*. 2001;42:1215–1225.

19. Margolis MK, Coyne K, Kennedy-Martin T, Baker T, Schein O, Revicki DA. Vision-specific instruments for the assessment of health-related quality of life and visual functioning: a literature review. *Pharmacoeconomics*. 2002;20:791–812.

20. Weih LM, Hassell JB, Keeffe J. Assessment of the impact of vision impairment. *Invest Ophthalmol Vis Sci*. 2002;43:927–935.

21. Klein BEK, Klein R, Jensen SC. A short questionnaire on visual function of older adults to supplement ophthalmic examination. *Am J Ophthalmol*. 2000;130:350–352.

22. Parrish RK, Gedde SJ, Scott IU, et al. Visual function and quality of life among patients with glaucoma. *Arch Ophthalmol*. 1997;115:1447–1455.

23. West SK, Munoz B, Rubin GS, et al. Function and visual impairment in a population-based study of older adults. The SEE project. *Invest Ophthalmol Vis Sci*. 1997;38:72–82.

24. Scott IU, Schein OD, West S, Bandeen-Roche K, Enger C, Folstein MF. Functional status and quality of life measurement among ophthalmic patients. *Arch Ophthalmol*. 1994;112:329–335.

25. Mangione CM, Gutierrez PR, Lowe G, Orav EJ, Seddon JM. Influence of age-related maculopathy on visual functioning and health-related quality of life. *Am J Ophthalmol*. 1999;128:45–53.

26. Stelmack J. Quality of life of low-vision patients and outcomes of low-vision rehabilitation. *Optom Vis Sci*. 2001;78:335–342.

27. Frost NA, Sparrow JM, Durant JS, Donovan JL, Peters TJ, Brookes ST. Development of a questionnaire for measurement of vision-related quality of life. *Ophthalmic Epidemiol*. 1998;5:185–210.

28. de Boer MR, Moll AC, de Vet HCW, Terwee CB, Völker-Dieben HJM, van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt*. 2004;24:257–273.

29. Vitale S, Schein OD. Qualitative research in functional vision. *Int Ophthalmol Clin*. 2003;43: 17–30.

30. Ehrlich JR, Spaeth GL, Carlozzi NE, Lee PP. Patient-centered outcome measures to assess functioning in randomized controlled trials of low-vision rehabilitation: a review. *Patient*. 2017;10:39–49.

31. Massof RW. A systems model for low vision rehabilitation. I. Basic concepts. *Optom Vis Sci*. 1995;72:725–736.

32. Massof RW. A systems model for low vision rehabilitation. II. Measurement of vision disabilities. *Optom Vis Sci*. 1998;75:349–373.

33. Keeffe JE, Lam D, Cheung A, Dinh T, McCarty CA. Impact of vision impairment on functioning. *Aust N Z J Ophthalmol*. 1998;26:S16–S18.

34. Keeffe JE, McCarty CA, Hassell JB, Gilbert AG. Description and measurement of handicap caused by vision impairment. *Aust N Z J Ophthalmol*. 1999;27:184–186.

35. Hassell J, Weih L, Keeffe J. A measure of handicap for low vision rehabilitation: the impact of vision impairment profile. *Clin Exp Ophthalmol*. 2000;28:156–161.

36. Massof RW, Ahmadian L, Grover LL, et al. The activity inventory: an adaptive visual function questionnaire. *Optom Vis Sci*. 2007;84:763–774.

37. Massof RW, Hsu CT, Baker FH, et al. Visual disability variables. I: the importance and difficulty of activity goals for a sample of low-vision patients. *Arch Phys Med Rehabil*. 2005;86:946–953.

38. Kirby RL, Basmajian J. The nature of disability and handicap. In: Kirby RL, Basmajian J, eds. *Medical Rehabilitation*. Baltimore: Willimans & Wilkins; 1984:14–18.

39. Massof RW. Likert and Guttman scaling of visual function rating scale questionnaires. *Ophthalmic Epidemiol*. 2004;11:381–399.

40. 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.

41. Mallinson T. Why measurement matters for measuring patient vision outcomes. *Optom Vis Sci*. 2007;84:E675–E682.

42. Massof RW, Ahmadian L. What do different visual function questionnaires measure? *Ophthalmic Epidemiol*. 2007;14:198–204.

43. Khadka J, McAlinden C, Pesudovs K. Quality assessment of ophthalmic questionnaires: review and recommendations. *Optom Vis Sci*. 2013;90:720–744.

44. Gothwal VK, Lovie-Kitchin JE, Nutheti R. The development of the LV Prasad-Functional Vision Questionnaire: a measure of functional vision performance of visually impaired children. *Invest Ophthalmol Vis Sci*. 2003;44:4131–4139.

45. Stelmack JA, Szlyk JP, Stelmack TR, et al. Psychometric properties of the Veterans Affairs low-vision visual functioning questionnaire. *Invest Ophthalmol Vis Sci*. 2004;45:3919–3928.

46. 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.

47. Geruschat DR, Bittner AK, Dagnelie G. Orientation and mobility assessment in retinal prosthetic clinical trials. *Optom Vis Sci*. 2012;89:1308–1315.

48. Massof RW. An interval-scaled scoring algorithm for visual function questionnaires. *Optom Vis Sci*. 2007;84:E690–E706.

49. Massof RW. Understanding Rasch and item response theory models: applications to the estimation and validation of interval latent trait measures from responses to rating scale questionnaires. *Ophthalmic Epidemiol*. 2011;18:1–19.

50. Humayun MS, Dorn JD, Da Cruz L, et al. Interim results from the international trial of Second Sight's visual prosthesis. *Ophthalmology*. 2012;119:779–788.

51. Goldstein JE, Massof RW, Deremeik JT, et al. Baseline traits of low vision patients served by private outpatient clinical centers in the United States. *Arch Ophthalmol*. 2012;130:1028–1037.

52. Coren S, Hakstian AR. Visual screening without the use of technical equipment: preliminary development of a behaviorly validated questionnaire. *Appl Opt*. 1987;26:1468–1472.

53. Andrich D. A rating formulation for ordered response categories. *Psychometrika*. 1978;43:561–573.

54. Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2001.

55. Stelmack J, Szlyk JP, Stelmack T, et al. Use of Rasch person - item map in exploratory data analysis: a clinical perspective. *J Rehabil Res Dev*. 2004;41:233–242.

56. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci*. 2007;84:663–674.

57. Linacre JM. Structure in Rasch residuals: why principal components analysis (PCA)? Rasch Measurement Transactions. 1998. Available at: http://www.rasch.org/rmt/rmt122m.htm. Accessed 1 April, 2017.

58. Calderon JL, Morales LS, Liu H, Hays RD. Variation in the readability of items within surveys. *Am J Med Qual*. 2006;21:49–56.

59. Mangione CM, Lee PP, Gutierrez PR, et al. Development of the 25-item National Eye Institute visual function questionnaire. *Arch Ophthalmol*. 2001;119:1050–1058.

60. Gothwal VK, Sumalini R, Bharani S, Reddy SP, Bagga DK. The second version of the L. V. Prasad-functional vision questionnaire. *Optom Vis Sci*. 2012;89:1601–1610.

61. Khadka J, Ryan B, Margrain TH, Court H, Woodhouse JM. Development of the 25-item Cardiff Visual ability auestionnaire for children (CVAQC). *Br J Ophthalmol*. 2010;94:730–735.

62. Dagnelie G, Jeter PE, Adeyemo O. Optimizing the ULV-VFQ for clinical use through item set reduction: psychometric properties and trade-offs. *Transl Vis Sci Technol*. In Press.

63. Dagnelie G, Barry MP, Adeyemo O, Jeter PE, Massof RW. Twenty Questions: an adaptive version of the PLoVR ultra-low vision (ULV) questionnaire. *Invest Ophthalmol Vis Sci*. 2015;56:497–497.