The Activities of Daily Vision Scale for Cataract Surgery Outcomes: Re-evaluating Validity with Rasch Analysis

Konrad Pesudovs,[1] *Estibaliz Garamendi,*[1] *John P. Keeves,*[2] *and David B. Elliott*[1]

PURPOSE. The Activities of Daily Vision Scale (ADVS) has been extensively validated by traditional methodology. In the current study, Rasch analysis was used to explore further the validity of the ADVS and to determine whether improvements could be made.

METHODS. Forty-three patients with cataract underwent visual acuity (VA) and contrast sensitivity (CS) testing and completed the ADVS. The data were Rasch analyzed and the value of response scale and item reduction explored. A shortened version and the original ADVS were tested for criterion validity by determining correlations with VA and CS.

RESULTS. The ADVS data contained nonnormally distributed items and items with ceiling effects and empty response categories. Therefore, items benefited from shortening the response scale, the optimum length being three responses. There was poor targeting of item difficulty to patient ability, because many patients with cataract were sufficiently able that they had no difficulty with many activities. Items were eliminated if the task was too easy or did not fit with the overall concept of visual disability determined by the Rasch model. A reduced ADVS version was established that had adequate precision, equivalent criterion validity, and improved targeting of item difficulty to patient ability, but this version was still not ideal.

CONCLUSIONS. Despite careful traditional validation, the ADVS data contained inadequacies exposed by Rasch analysis. Through Rasch scaling, particularly with response scale reduction, the ADVS can be improved, but additional questions seem to be needed to suit the more able, including patients undergoing second eye cataract surgery. There remains a need to develop Rasch-scaled measures of visual disability for use in ophthalmic outcomes research. (*Invest Ophthalmol Vis Sci.* 2003;44:2892–2899) DOI:10.1167/iovs.02-1075

In recent years, measurement of patient-centered outcome has become accepted as an essential part of outcomes research. Such measures usually take the form of questionnaires, often considered to be measures of quality of life or disability. In ophthalmic research, these outcomes investigations have predominantly involved the measurement of visual disability. Numerous questionnaires exist for this purpose: the Visual Functioning Index (VFI),[1] the Visual Activities Questionnaire (VAQ),[2] the Activities of Daily Vision Scale (ADVS),[3] the Visual Performance Questionnaire (VPQ),[4] the 14-item Visual Functioning Index (VF-14),[5] and the Visual Disability Assessment (VDA),[6] among others.[7,8] These have been used extensively for studies of treatment outcome, especially for cataract surgery.[9–18]

Questionnaire scores provide a simple and convenient numerical representation of patient-centered outcome. The overall score is usually arrived at by adding up ordinal numerical values assigned to the subject's ratings for a series of questions. Such a response scale is called a Likert scale, and simply adding up the scores is called Likert scoring.[19] However, the validity of such scores has been called into question by modern test theory, which includes Rasch analysis.[20–23] At issue is what justification exists for assigning numerical values to responses and how a series of scores should be added together to produce an overall score. Although Likert scaling assumes all questions are equally weighted, Rasch analysis assumes items vary in difficulty. Rasch analysis calculates item difficulty in relation to person ability and weights overall scores accordingly. The scores are on a linear scale, allowing easy comparison of measures. Other benefits of Rasch analysis include powerful investigation of instrument validity particularly the fit of items to the overall construct, and the effectiveness of targeting of items to patients.

In recent times, Rasch analysis has been applied to disability measurement for both the validation or modification of existing scales or the development of new scales in many areas of medicine, including rheumatology,[24] rehabilitation medicine,[25] gerontology,[26] and overall health-related quality of life.[27] It has been applied to measures of visual disability in low-vision populations[28–30,31] and has been used to examine the validity of one visual disability questionnaire suitable for a cataract population: The 14-item Visual Functioning Index (VF-14).[23,32]

We have chosen the Activities of Daily Vision Scale (ADVS) as the subject of our investigation, because it was the first widely used and (traditionally) validated visual disability questionnaire and remains one of the most commonly used.[3,6,8,14,33–39] Much of what is accepted about patient-centered outcomes of cataract surgery relies on the validity of the ADVS and VF-14. Therefore, we think it is important to examine the psychometric properties of the ADVS in a Rasch model.

METHODS

Patients

The data used for this study were from a series of patients awaiting cataract surgery and were collected for a previously published study.[16] Forty-three patients (mean age, 72.6 ± 8.1 years) were included. Eighteen had bilateral cataract and 25 had one pseudophakic eye and were awaiting second eye surgery. No patients had comorbid eye disease. Informed consent was obtained from all subjects after the nature of the study had been fully explained. The tenets of the Declaration of Helsinki were followed, and the study gained approval from the university's ethics committee.

TABLE 1. Assessment of Item Quality in ADVS Data from 43 Patients with Cataract

Item	Description	Skew	Kurtosis	Missing Data (%)	Ceiling Effect (%)	Infit Mean Square	Outfit Mean Square	Item Calibration (SE)
15-Item scale								
1a–c	Driving at night	−1.00	0.38	26	19	0.77	0.76	63.8 (2.2)
1d	Seeing moving objects with night driving	−1.04	0.37	33	35	0.92	0.82	56.9 (2.6)
1e	Oncoming headlights	−0.29	−0.87	33	19	1.05	1.13	65.4 (2.2)
2a–c	Daytime driving	−2.39	5.09	19	63	1.37	1.35	45.2 (3.0)
2e	Drive in unfamiliar areas	−1.44	1.14	40	37	0.87	0.70	54.5 (2.8)
3a–c	Read signs at night	−0.22	0.84	14	16	0.98	0.98	67.0 (1.9)
4a–c	Read signs during the day	−1.78	2.16	2	65	1.54	1.03	51.1 (2.4)
8a–c	See/recognize faces	0.62	−0.29	5	23	1.07	1.08	62.9 (1.9)
9a–c	See television	−1.88	2.80	0	77	0.97	0.54	38.9 (3.3)
10a–c	Read writing on television	−1.08	0.25	2	47	0.95	1.01	54.2 (2.2)
11a–c	Read newspapers	−2.05	3.66	2	63	1.50	1.95	50.2 (2.4)
12a–c	Read medicine bottles	−0.81	−0.47	5	37	1.02	1.06	61.2 (2.0)
13a–c	Read food cans	−1.70	4.09	16	42	0.67	0.80	50.4 (2.5)
14a–c	Write checks	−3.78	15.52	2	86	2.02	0.84	36.6 (3.5)
15a–c	Thread a needle	−0.99	1.31	35	19	0.78	0.82	59.7 (2.4)
22-Item scale								
5a–c	Use public transport	−3.14	10.13	72	23	1.78	0.80	44.1 (5.0)
6a–c	Walk downstairs in daylight	−1.94	3.63	49	30	2.18	1.34	48.1 (3.3)
7a–c	Walk downstairs in dim light	−1.09	0.11	51	26	2.17	1.62	55.4 (2.9)
16a–c	Use a ruler, yardstick, or tape measure	−3.60	13.14	16	74	0.72	0.27	34.2 (4.2)
17a–c	Use a screwdriver	−3.12	10.06	23	65	0.59	0.29	37.7 (4.0)
18a–c	Prepare meals	−4.18	19.36	16	74	0.69	0.66	31.4 (4.5)
19a–c	Play cards	−3.22	10.48	40	53	0.32	0.14	30.9 (5.6)

Data are from Elliott et al.[16] Values of skew, kurtosis, percentage of missing data, and percentage ceiling effect (percentage answers in the no difficulty response) are shown. The Rasch analysis fit statistics (infit and outfit mean squares), item calibration, and standard error are given for the 22-item, five-response model, although items are sorted into those retained in a reduced 15-item version and that are additional in the original 22-item version.

Clinical Assessment

Logarithm of the minimum angle of resolution (LogMAR) visual acuity (VA) was measured monocularly and binocularly using standard Early-Treatment Diabetic Retinopathy Study (ETDRS) charts at 4 m with a luminance of 100 cd/m^2 and letter-by-letter scoring, and contrast sensitivity (CS) was measured with the Pelli-Robson chart at 3 m with a luminance of 100 cd/m^2 and letter-by-letter scoring.[16] Only the binocular VA and CS results are included in the analyses for this study, because these have been shown to be more closely related to disability than better or worse eye measurements.[40,41]

Disability Assessment

Visual disability was assessed with the Activities of Daily Vision Scale (ADVS). The original ADVS contains 22 items, each of which examine the patient's ability to perform an activity. These activities are listed in Table 1. The 22 items are examined with 61 questions, usually 3 questions per item. The first is to assess whether the patient engages in the activity (if not it is "Not Applicable" which is treated as missing data), the second scales patient responses from no difficulty (5), to a little difficulty (4), moderate difficulty (3), and extreme difficulty (2). The third question asks whether the patient is unable to perform the activity because of poor vision (if not, it is missing data; if so then the most disabled score, 1, is assigned). Thus, in Table 1, many items are assigned three questions: for example, 1a–c covers driving at night. For the convenience of discussion and Rasch analysis, scores from such groups of three questions pertaining to the same item are considered to fall on one scale, which is consistent with the original ADVS scoring system.[3]

Analysis

The responses of the 43 patients to the items of the ADVS were examined in terms of distribution, Cronbach α, and Rasch Analysis. The distribution of responses was examined for compliance with normality (skew and kurtosis), missing data, and ceiling effect (the percentage of responses in the most able end category of the response scale). Various versions of the ADVS were compared with the original ADVS with criterion validity testing by comparing the visual measures of VA and CS with disability scores, using Spearman rank correlations. These analyses were performed on computer (SPSS for Windows; SPSS Sciences, Chicago, IL).

Rasch analysis was also performed on computer (Winsteps ver. 3.35, produced by John M. Linacre,[42] which calculates the Wright and Masters[43] version of Rasch model estimates, by using joint maximum likelihood estimation. In using Rasch analysis, we implicitly assume as a goal reengineering an assessment so that it is better targeted to the people who will be evaluated with it. Targeting implies that the challenge presented by the items and the response categories of the rating scale cover the same range of abilities as are found in the population to be measured. The first step in improving targeting of items to patients was to consider whether the response scale categories were used appropriately across the whole test. If this was not the case then the effect of merging response scale categories was investigated. The third step was to consider the value of removing items from the questionnaire if they were not effective in contributing to the measurement of the abilities of persons of interest.

RESULTS

The clinical, visual, and disability results for this series of patients have been published.[16] VA in the surgically treated eye for the bilateral cataract cases was 0.54 ± 0.36 logMAR and for the second eye cases was 0.52 ± 0.29 logMAR (Snellen equivalents approximately 20/60). Binocular VA in the bilateral cataract cases was 0.18 ± 0.18 logMAR (Snellen equivalent 20/30) and in the second eye cases was 0.05 ± 0.11 logMAR (Snellen equivalent 20/20^{-2}).

```
<more able>|<more difficult>
100      X  +
            |
            |
          T|
            |
         X  |
 90         +
        XX  |
            |
        XX  |
       XXX S|
        XX  |
 80      X  +
      XXXX  |
      XXXX  |
        XX  |
         X  |
         X  M|T
 70   XXXX  +
        XX  |
            |  read signs at night
         X  |  oncoming headlights
            |  driving at night        see faces
        XX  |S read medicine bottle
 60    XXX  +  thread needle
           S|
         X  |  moving objects & night driving
        XX  |  drive unfamiliar areas     steps at night        teletext
         X  |
            |  read signs daytime
 50      X  +M read food cans        read newspapers
            |  steps daytime
         X T|
            |  daytime driving
            |  public transport
            |
 40         +
            |S television           use screwdriver
            |  write checks
            |  tape measure
         X  |
            |  play cards            prepare meals
 30         +
   <less able>|<less difficult>
```
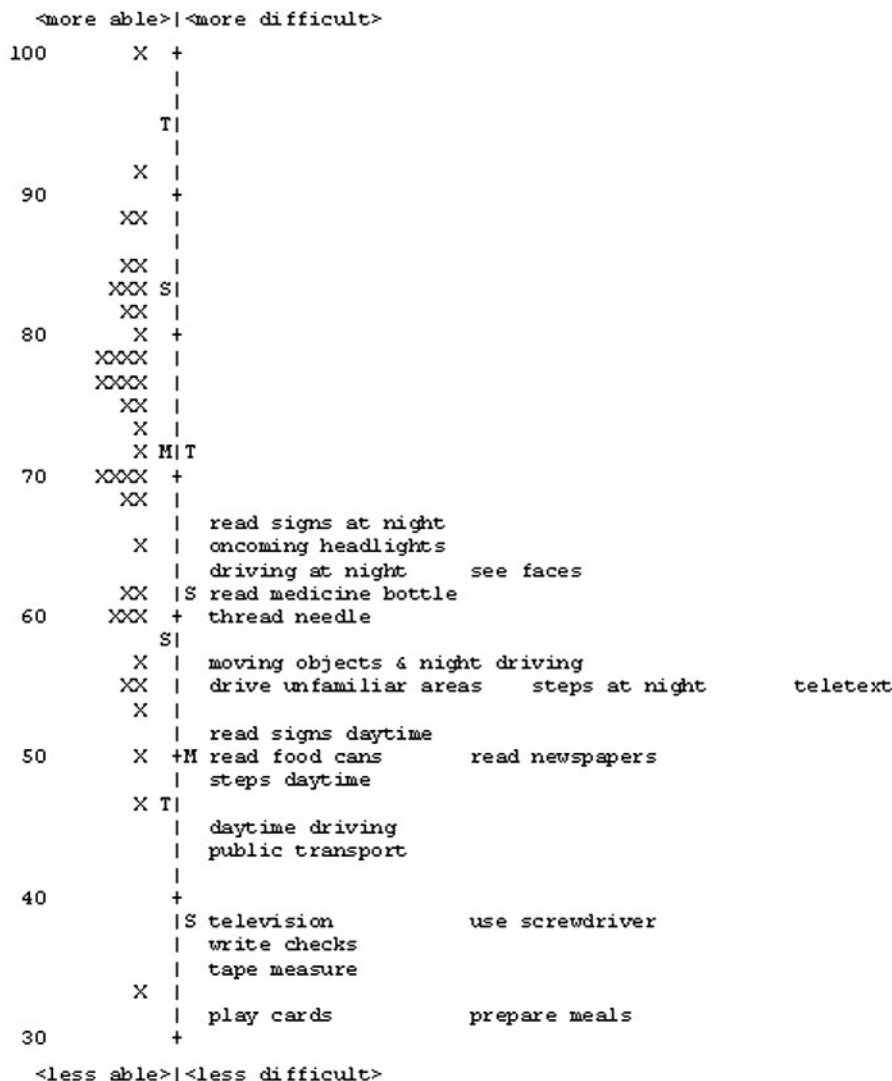
FIGURE 1. Patient ability/item difficulty map for the 22-item ADVS. To the *left* of the *dashed line* are the patients, represented by X, and on the *right* are the items denoted by their content. More able patients and more difficult items are near the *top* of the diagram with less able patients and easier items near the *bottom*. The scale is in units (0–100). M, mean; S, 1 SD from the mean; T, 2 SD from the mean.

Descriptive Statistics

The 22 items of the ADVS are listed in Table 1 along with several criteria for assessing the quality of these items: compliance with normality (skew and kurtosis), ceiling effect (percentage of responses in the most able end category of the response scale), and the percentage of cases with missing data. The data from these 43 patients suggest that 12 items do not provide normally distributed data, if normality is defined as skew and kurtosis within the range −2.00 to +2.00.

Rasch Analysis

Figure 1 shows a patient ability/item difficulty map determined by Rasch analysis for the original 22-item ADVS. Patients (Xs on the left) appear in ascending order of ability from the bottom of the map to the top, and items (item names on the right) appear in ascending order of difficulty from the bottom to the top. Both patients and items appear along the same scale, which in this case is a linear transformation of the Rasch logit scale to fit a 0 to 100 scale (Winsteps Uscale = 8.10). In this data set, the items are, on the whole, too easy for the abilities of the patients, which is represented by the X's located higher and item names located lower. This illustrates poor targeting of item difficulty to patient ability. It may appear that there are no items to discriminate between the more able patients (X's not opposed by any items) at the top of the map. However, Figure 1 presents the item calibrations averaged across the 5-point rating scale. If each step of the rating scale were illustrated, some of these more able patients would be shown to be targeted by the thresholds between no difficulty and a little difficulty for the more challenging items. However, four patients were beyond all steps for all items, and therefore differences in their visual abilities are poorly discriminated by the ADVS. There is a floor effect at the bottom of the map where there are many items targeting few patients (items are too easy). If the items were well targeted to the patients, the means of the two distributions, denoted in Figure 1 by M, would be close to each other. We attempted to address the problems highlighted in Figure 1 through response-scale reduction and item reduction.

Response-Scale Reduction

For 11 of the 22 items the "don't do this activity because of visual problems" category was not used by any of the 43 patients (for four of these items, the "extreme difficulty" category was also not used). For the 11 items in which there were responses in the end category, the number varied from one (2.3%) to three (7.0%). This underutilization was overcome by combining response 1 "don't do this activity because of visual

TABLE 2. Person Separation, Targeting of Items to Patients, Cronbach α, and Criterion Validity for the Original ADVS and the Four Rasch–scaled Versions

	22-Item ADVS Not Rasch Scaled	22 Item ADVS (5 Responses)	22 Item ADVS (4 Responses)	22 Item ADVS (3 Responses)	15 Item ADVS (3 Responses)
Person Separation	—	2.37	2.48	2.53	2.22
Difference between item and patient means (patient SEM)	—	20.96 (1.94)	16.85 (2.12)	12.54 (2.41)	6.95 (2.34)
Cronbach α	0.92	0.92	0.92	0.92	0.91
Correlation with VA	−0.43	−0.44	−0.41	−0.43	−0.44
Correlation with CS	0.45	0.50	0.46	0.47	0.48

problems" and response 2 "extreme difficulty." This modestly improved targeting of item difficulty to patient ability, illustrated by a reduction in the difference between the mean value for the patients and the mean value of the items from 20.96 units to 16.85 units (Table 2).

The combined-response category was still empty for four items. For the remaining 18 items, the category's content varied from 1 (2.3%) to 10 (23.3%) patients. However, only one question had more than 20% in the combined-response category, and only four others had more than 10%. More than half of the items had less than 5% of the data in this combined-response category. Therefore, it was decided to combine the response category "moderate difficulty" with the already combined "don't do this activity because of visual problems" and "extreme difficulty." This further reduced the difference between the mean value for the patients and the mean value of the items to 12.54 units (Table 2). Response scale reduction also succeeded in reducing skew and kurtosis for 10 of the items.

Item Reduction

Rasch analysis was used to improve the ADVS by improving internal consistency (item fit to the model) and improving targeting of the test to the population (reducing redundant or underutilized items). The Rasch model fit statistics, infit and outfit mean square, which compare the predicted responses to those observed, were used to monitor the compatibility of the data with the model. Outfit (outlier-sensitive fit) mean square is the conventional sum of squared standardized residuals and is sensitive to occasional responses that are very different from the expected response. For infit (information-weighted fit) mean square, each squared standardized residual value is first weighted by its variance and then summed. In this way, infit takes less notice of extreme responses as it is weighted to be sensitive to responses that are close to a respondent's level of (in this case) visual function. Both infit and outfit mean squares have an expected value of 1, with those less than 0.80 representing items that overfit the model and are too predictable (they have at least 20% less variation than was expected). Overfitting items may be redundant or noncontributory, because they lack variance. Mean squares greater than 1.20 represent misfit (at least 20% more variance than was expected). A high item infit or outfit suggests that the item measures something different than the overall scale.[44] A high item outfit may also indicate that an item is influenced by, in this case, visual disability in some patients with cataract but not in all cases. Such items would be acceptable, as long as they are not too extreme. Based on this rationale, we determined that infit mean square should drive item reduction, so more stringent criteria were used for infit and more lenient for outfit. The criteria (in order of priority) used to identify candidate items for removal were: (1) infit mean square outside 0.80 to 1.20; (2) outfit mean square outside 0.70 to 1.30; (3) high proportion

of missing data; (4) ceiling effect—a high proportion reporting no difficulty; and (5) skew and kurtosis outside −2.00 to +2.00.

Patient fit to the model was checked before considering items for removal to identify whether abnormally fitting patients may be contributing to abnormally fitting items. High item infit and outfit responses could be due to rogue responses from a small number of patients (unlike low values—i.e., responses that are too predictable—which are produced by most of the patients). Six patients had infit greater than 1.40. All patients of high infit tended to have no trouble with most items and so much trouble they could not perform one item. This was not a typical response pattern, as most patients who had difficulty with one item had difficulty with multiple items. None of the items that caused these patients difficulty were items with high infit. Therefore, none of these cases of unusual response patterns were driving items to misfit, and all patients were therefore retained in the analysis for the evaluation of items. The adequacy of the sample size was shown by the average SE of the persons (5.0 units) being about half the average size of a rating scale step (9.9 units).

Item reduction was an iterative process, one item removed at a time, fit to the model reestimated accordingly (fit is relative, so removal of items leads to changes in fit). The item with the highest number of candidate criteria, ordered by priority, was removed first. The infit and outfit mean squares for all 22 items are shown in Table 1. The four items at the bottom of Table 1 (19a–c, play cards; 18a–c, prepare meals; 17a–c, use a screwdriver; and 16a–c, use a ruler, yardstick, or tape measure) have low infit mean squares, high skew, kurtosis, and ceiling effect and are seen at the bottom of the map in Figure 1 in a group of six items with mean values that aligned with just one patient. These four items were so easy that most patients could perform them without difficulty, and they were the first four items removed. The next item with the poorest fit statistics was 5a–c, use public transport (infit 0.74, outfit 0.39), which had misfitted initially (infit 1.78) before the removal of the first four items. This volatility was probably related to the 72% missing data, and so this item was removed also. The removal of items with low infit improved the fit of several of the items that started with high infit: for example, 2a–c, daytime driving (infit became 1.02), and 11a–c, reading newspapers (infit became 0.96). However, two items still had high infit mean squares (and 50% missing data) and so were removed (6a–c, walk downstairs in daylight, infit 1.38; 7a–c, walk downstairs in dim light, infit 1.34). This resulted in a 15-item ADVS with a person separation of 2.22. Person separation is an indication of the precision with which the variability present in the patients is captured by the test, expressed as the ratio of the adjusted SD to the root mean square error. Therefore a higher patient separation (>2.0) is indicative that patients are significantly different in ability across the measurement distribution. However, this questionnaire still contains a poorly

fitting item (14a–c, writing checks: infit 0.77, outfit 0.26) with a large ceiling effect, skew, and kurtosis. Its removal led to another poorly fitting item. Item removal could be continued until all items fit well, and this occurred with 11-items remaining. However, person separation was decreased to 1.61, so it was decided to retain a poorly fitting item in a 15-item questionnaire, rather than lose person separation (precision). These items could be removed and replaced with other items of a difficulty level that better targets patient ability if such items were available. The original validation of the ADVS found a Cronbach α of between 0.91 and 0.94[3]; in this data set it was 0.92. Cronbach α was unchanged at 0.91 for the shortened version (Table 2).

Criterion Validity

The Spearman rank correlation of VA and CS with the original 22-item ADVS score was −0.43 and 0.45. This was not changed by Rasch scaling alone or by response scale or item reduction (Table 2).

DISCUSSION

Although the data set used in this study was relatively small, it was large enough to highlight the inadequacies of traditional assessment and validation of questionnaires developed to measure visual disability. The fundamental weakness of Likert scoring is that all items are assumed to be equally difficult and all scores of the same worth. The illogic of this can be illustrated with the ADVS item pairs of driving during the day and driving at night, or reading food cans and reading medicine bottles. As most people would expect,[23] driving at night in Figure 1 was more difficult than driving during the day, and reading medicine bottles was more difficult than reading food cans. With Rasch scaling, the difficulty of these items weights the score—for example, a patient with a little difficulty with driving at night would be expected to have an overall score of approximately 62 units (scale, 0–100), whereas a patient with a little difficulty doing the easier task of driving during the day would be expected to have a poorer overall score of approximately 48 units. In this way, the overall score includes a contribution from each item weighted by its difficulty.

The linear scale allows easy comparison of this relative difficulty of items and relative ability of patients. Moreover, because items and persons are measured on the same scale, the targeting of item difficulty to patient ability is readily illustrated. Reading food cans is found at the calibration point for the mean of the item group (Fig. 1), whereas the mean of the patient group is a long way farther up the scale. This illustrates poor targeting of item difficulty to patient ability. Many of the items are too easy for this cataract population to be troubled by (e.g., preparing a meal, playing cards, or using public transport), and so the visual disability of these patients is poorly measured. Combining the response categories that cater to the most disabled cases improved the targeting of item difficulty to patient ability. This weights the scale toward the less disabled end, thus giving it more power to discriminate between the more able patients. This appears to be a clinically sensible change,[45] with the shortened version of the ADVS just having to discriminate between patients with "no difficulty," "a little difficulty," and "at least a moderate amount of difficulty." Response scale reduction has also decreased skew and kurtosis for most of the items.[46] Although normal data are not a prerequisite for Rasch analysis,[45,47] items that are normally distributed are more likely to contribute to person separation than skewed or kurtotic items. Person separation increases, although not significantly, from 2.37 with a 5-point response scale to 2.53 with a 3-point response scale.

Some of the 22 items in the original ADVS provided poor data due to missing data, ceiling effects, or poor Rasch fit statistics. The misfit for the two walking-downstairs questions may be due to high amounts of missing data (Table 1) and/or may be because these were the only mobility questions. Although mobility is an important component of visual ability, Stelmack et al. have shown that, in a low-vision population, mobility tasks do not tend to fit well with reading tasks in a Rasch model (Stelmack J, Szlyk J, Stelmack T, Ardickas Z, Massof R, ARVO Abstract 3816, 2002). Given that the fit to a Rasch model is in part a function of the items sharing similar content, which in the ADVS is chiefly driving and reading, it is not surprising that two mobility items fit poorly. The 15-item version has one poorly fitting item (14a–c writing checks) although when this is removed another fits poorly. Although these items could be removed, any benefit is counteracted by a decrease in person separation. This could be overcome if new items that were relevant to most patients with cataract and of greater visual difficulty, were added.

Our results are similar to those of Velozo et al.[32] who looked at the validity of the VF-14, another traditionally validated visual disability questionnaire, in terms of a Rasch model. They too found some poorly fitting items, poor discrimination, redundancy and underutilization of response categories. They found that the VF-14's 5-category response scale could be reduced to a three-category response scale. Again, it was the categories representing the greatest difficulty that were sparingly used and could be combined. They found that the redundancy could be reduced through item removal without loss of internal consistency, but the VF-14 discrimination and redundancy problems could not be solved by response scale and item reduction alone. Velozo et al. added 10 extra items to try to improve discrimination. This was only partly effective, but they produced a VF-10 with psychometric properties superior to those of the VF-14. As all other visual disability questionnaires (e.g., VFI, VAQ, VPQ, and VDA) that have been developed use Likert scoring methods,[23] it is likely that they all have the same problems, at least when used to assess visual disability in patients undergoing cataract surgery. Although Rasch scaling could be used to examine the data collected on existing visual disability questionnaires, it is likely that optimal validity will be achieved through further modification of these questionnaires or the development of new questionnaires.[23] Rasch analysis has only been used in the development of three visual disability questionnaires, and these have been for low-vision populations.[28–30] It is unlikely that these would be suitable for cataract outcomes research, because cataract populations are likely to be much less impaired.

The poor targeting of item difficulty to patient ability in the ADVS raises the question of whether our cataract population was typically or less impaired than average. The mean VA in the surgically treated eyes in this study is comparable to that in many other current series,[18,48] including unilateral cataract series.[49–51] The binocular VA is similar to that in series with mixed first eye and second eye surgeries,[9,11,12,33,52] but better than that seen in bilateral cataract only series.[18] Similarly, the VA in this series is better than that with comorbidity,[17,52,53] and that in British patients on waiting lists in the United Kingdom.[54] VA is also better in this series than in older series.[55,56] It is well known that there have been changing indications for cataract surgery, due to the increased efficiency and safety of the procedure, so that it is now offered at a lower level of impairment.[57] This suggests that whereas the ADVS may have been ideal when it was being developed in the late 1980s and early 1990s, it is no longer suited to the more visually able patients who undergo surgery today. It also suggests that the ADVS may be more suited to measuring disability in bilateral cataract and perhaps in cases with comorbidity. To
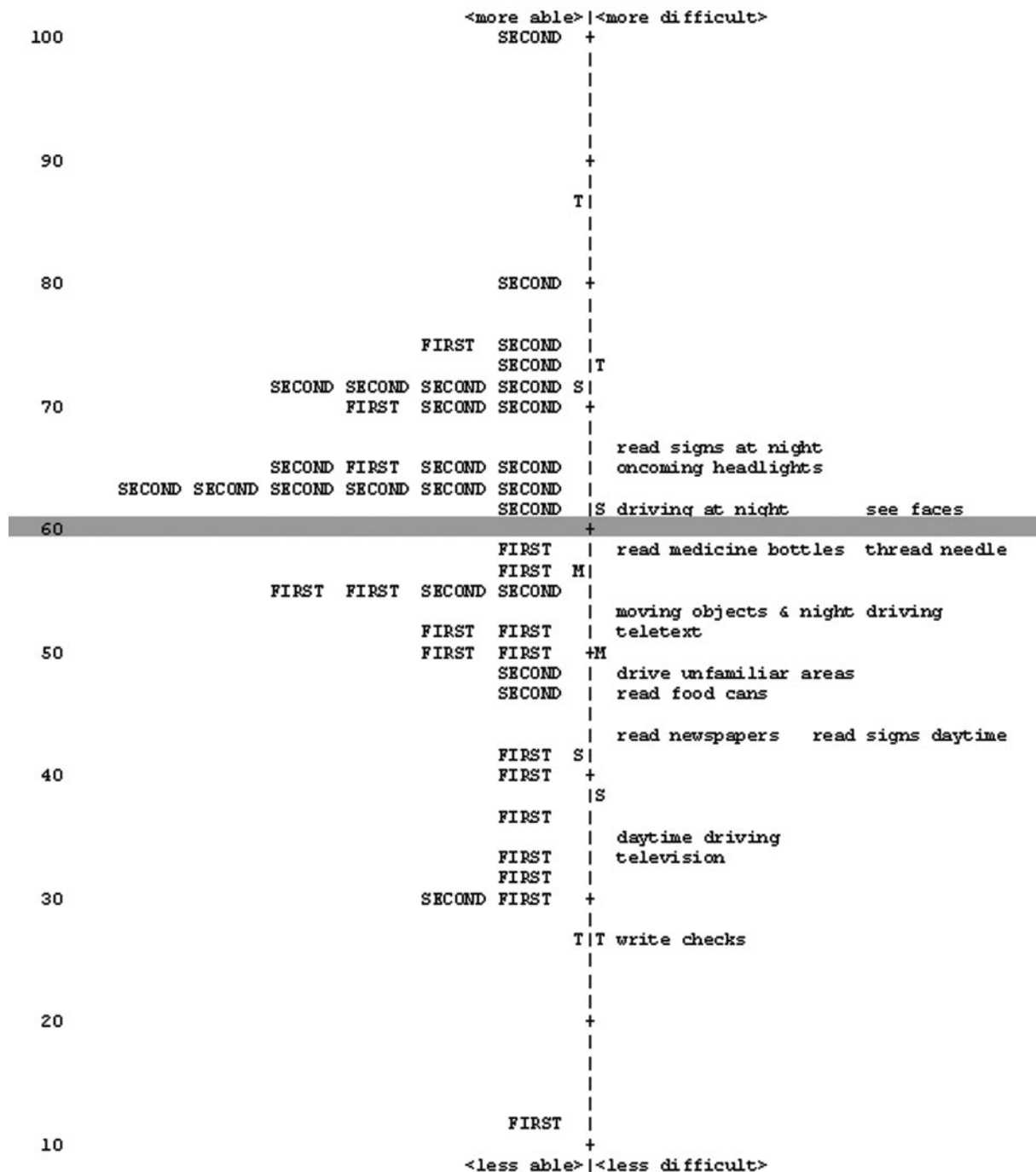
```
                          <more able>|<more difficult>
  100                          SECOND  +
                                      |
                                      |
                                      |
                                      |
   90                                 +
                                      |
                                    T |
                                      |
                                      |
   80                          SECOND  +
                                      |
                                      |
                      FIRST   SECOND  |
                               SECOND |T
        SECOND SECOND SECOND   SECOND S|
   70           FIRST  SECOND  SECOND  +
                                      |
                                      |  read signs at night
                SECOND FIRST   SECOND SECOND |  oncoming headlights
      SECOND SECOND SECOND SECOND SECOND SECOND |
                               SECOND |S driving at night       see faces
   60 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━+━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━
                                FIRST |  read medicine bottles  thread needle
                                FIRST M|
        FIRST  FIRST  SECOND   SECOND |
                                      |  moving objects & night driving
                         FIRST  FIRST |  teletext
   50                    FIRST  FIRST +M
                               SECOND |  drive unfamiliar areas
                               SECOND |  read food cans
                                      |
                                      |  read newspapers    read signs daytime
                                FIRST S|
   40                           FIRST  +
                                      |S
                                FIRST |
                                      |  daytime driving
                                FIRST |  television
                                FIRST |
   30                    SECOND FIRST  +
                                      |
                                    T |T write checks
                                      |
                                      |
                                      |
   20                                 +
                                      |
                                      |
                                      |
                                      |
                                      |
                                FIRST |
   10                                 +
                          <less able>|<less difficult>
```

FIGURE 2. This patient ability/item difficulty map is for the 15-item ADVS, but it shows patients classified as waiting for either first eye (bilateral cataract) cataract surgery or second eye cataract surgery. *Above* the *shaded line* at 60 units are 23 patients, of whom 20 were awaiting second eye surgery and 3 first eye surgery. *Below* the *line* are 20 patients, of whom 16 were awaiting first eye surgery and 4 second eye surgery.

look at the importance of first eye surgery and second eye surgery in the ranking of ability by Rasch analysis, Figure 2 shows which patients were to undergo first eye surgery and which were to have second eye surgery. It can be seen that most of the second eye patients were more able and most of the first eye patients were less able than the average patient. This suggests that the ADVS may be more suitable for patients with bilateral cataracts and less for those needing second eye surgery. However, this is a problem for outcome studies, because after first eye cataract surgery, patients are prospects as preoperative second eye surgery cases. Moreover, the ADVS

and the VF-14 have been extensively used to look at the relative benefit of first and second eye cataract surgery.[9,12,16,52] Also, after second eye surgery poor targeting of items to patients is an even greater problem. Therefore, new disability scales are needed that can accurately measure visual disability in these groups. Perhaps questionnaires for patients with cataract should also contain items that tap issues of relevance to patients with unilateral visual loss (e.g., stereopsis, anisometropia, and inhibition[16,50,56]) and possibly should include domains of quality of life other than visual disability. This raises the possibility that separate questionnaires may be needed, be-

cause the items relevant to patients with binocular visual loss may be different from those relevant to patients with unilateral loss.

Cronbach α

We found the original version of the ADVS to have a Cronbach α of 0.92, which is comparable to the original validation.[3] This is considered to be exceptionally high and may be indicative of redundancy.[58] Indeed, Rasch analysis highlights redundancy within our ADVS data, which was eliminated through removal of items for the 15-item version. Redundancy is a problem if the process of creating the overall score for the questionnaire involves just adding all the item scores together. In such a case, the overall score overweights the importance of the issue that is served by redundant items.[59] It is possible to have a high Cronbach α through inclusion of items that are highly correlated. Similarly, because Cronbach α is not independent of the number of items, it may be elevated by including many items. Furthermore, Cortina[60] has shown that any test of 20 items would have a high Cronbach α, and its use in this case is therefore not particularly helpful. For these reasons, Cronbach α should probably be considered to be more of a traditional measure than a useful measure.[23]

Criterion Validity

All four reduced versions also show correlations with VA and CS that are at least as good as those found with the original 22-item version. For the 15-item version, the failure of the removal of 7 items to damage the relationship between vision and ability confirms the criterion validity of the shortened version. Moreover, it is not possible to achieve high correlations with VA if measures fall across only a narrow range, as ours did.[61]

CONCLUSIONS

We have shown that Rasch analysis can highlight the inadequacies of the traditional method for questionnaire development and validation. Reducing the number of items in the ADVS while maintaining credible psychometric properties is an advantage in terms of efficiency. The shorter questionnaire should be easier and faster to use and, as such, is more likely to be filled out accurately.[62] Nevertheless, the 15-item scale is not an ideal scale. There remains poor targeting of item difficulty to patient ability, which means that patients with cataract are still too able for the items. The reduced ADVS could be refined further through item reduction, but this unacceptably reduces person separation, and therefore new items are needed that represent tasks that are more difficult, in line with patients' ability. There is still a need to develop Rasch-scaled visual disability questionnaires for use in surgical outcomes assessment, research, clinical care, and waiting-list prioritization that show good validity and test–retest repeatability.

References

1. Bernth-Petersen P. Visual functioning in cataract patients: methods of measuring and results. *Acta Ophthalmol*. 1981;59:198–205.
2. Sloane ME, Ball K, Owsley C, Bruni SR, Roenkar DL. The visual activities questionnaire: developing an instrument for assessing problems in everyday visual tasks. *Tech Dig Noninvas Assess Vis Sys*. 1992;1:26–29.
3. Mangione CM, Phillips RS, Seddon JM, et al. Development of the "Activities of Daily Vision Scale": a measure of visual functional status. *Med Care*. 1992;30:1111–1126.
4. Bergman B, Sjostrand J. Vision and visual disability in the daily life of a representative population sample aged 82 years. *Acta Ophthalmol*. 1992;70:33–43.
5. Steinberg EP, Tielsch JM, Schein OD, et al. The VF-14: an index of functional impairment in patients with cataract. *Arch Ophthalmol*. 1994;112:630–638.
6. Pesudovs K, Coster DJ. Validation of a new tool for the assessment of subjective visual disability. *Br J Ophthalmol*. 1998;82:617–624.
7. Massof RW, Rubin GS. Visual function assessment questionnaires. *Surv Ophthalmol*. 2001;45:531–548.
8. 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.
9. Javitt JC, Brenner MH, Curbow B, Legro MW, Street DA. Outcomes of cataract surgery: improvement in visual acuity and subjective visual function after surgery in the first, second, and both eyes. *Arch Ophthalmol*. 1993;111:686–691.
10. Desai P, Reidy A, Minassian DC, Vafidis G, Bolger J. Gains from cataract surgery: visual function and quality of life. *Br J Ophthalmol*. 1996;80:868–873.
11. Pesudovs K, Coster DJ. Cataract surgery reduces subjective visual disability. *Aust NZ J Ophthalmol*. 1997;25:3–5.
12. Elliott DB, Patla A, Bullimore MA. Improvements in clinical and functional vision and perceived visual disability after first and second eye cataract surgery. *Br J Ophthalmol*. 1997;81:889–895.
13. Uusitalo RJ, Brans T, Pessi T, Tarkkanen A. Evaluating cataract surgery gains by assessing patients' quality of life using the VF-7. *J Cataract Refract Surg*. 1999;25:989–994.
14. Superstein R, Boyaner D, Overbury O. Functional complaints, visual acuity, spatial contrast sensitivity, and glare disability in preoperative and postoperative cataract patients. *J Cataract Refract Surg*. 1999;25:575–581.
15. Castells X, Alonso J, Ribó C, et al. Comparison of the results of first and second cataract eye surgery. *Ophthalmology*. 1999;106:676–682.
16. Elliott DB, Patla AE, Furniss M, Adkin A. Improvements in clinical and functional vision and quality of life after second eye cataract surgery. *Optom Vis Sci*. 2000;77:13–24.
17. Mallah MK, Hart PM, McClure M, et al. Improvements in measures of vision and self-reported visual function after cataract extraction in patients with late-stage age-related maculopathy. *Optom Vis Sci*. 2001;78:683–688.
18. Saw SM, Tseng P, Chan WK, Chan TK, Ong SG, Tan D. Visual function and outcomes after cataract surgery in a Singapore population. *J Cataract Refract Surg*. 2002;28:445–453.
19. Likert RA. A technique for the measurement of attitudes. *Arch Psychol*. 1932;140:1–55.
20. Wright BD, Linacre JM. Observations are always ordinal, measurements must be interval. *Arch Phys Med Rehabil*. 1989;70:857–860.
21. Fisher WP Jr. The Rasch debate: validity and revolution in educational measurement. In: Wilson M, ed. *Objective Measurement: Theory into Practice*. Norwood, NJ: Ablex; 1994:36–72.
22. Fisher WP Jr, Eubanks R, Marier RL. Equating the MOS SF36 and the LSU HSI Physical Functioning Scales. *J Outcome Meas*. 1997;1:329–362.
23. Massof RW. The measurement of vision disability. *Optom Vis Sci*. 2002;79:516–552.
24. Eyres S, Tennant A, Kay L, Waxman R, Helliwell PS. Measuring disability in ankylosing spondylitis: comparison of bath ankylosing spondylitis functional index with revised Leeds Disability Questionnaire. *J Rheumatol*. 2002;29:979–986.
25. Hart DL, Wright BD. Development of an index of physical functional health status in rehabilitation. *Arch Phys Med Rehabil*. 2002;83:655–665.
26. Jette AM, Haley SM, Coster WJ, et al. Late life function and disability instrument: I. Development and evaluation of the disability component. *J Gerontol A Biol Sci Med Sci*. 2002;57:M209–M216.
27. Prieto L, Alonso J, Ferrer M, Anto JM. Are results of the SF-36 health survey and the Nottingham Health Profile similar?—a comparison in COPD patients. Quality of Life in COPD Study Group. *J Clin Epidemiol*. 1997;50:463–473.
28. Becker SW, Lambert RW, Schulz EM, Wright BD, Burnet DL. An instrument to measure the activity level of the blind. *Int J Rehabil Res*. 1985;8:415–424.

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

30. Turano KA, Geruschat DR, Stahl JW, Massof RW. Perceived visual ability for independent mobility in persons with retinitis pigmentosa. *Invest Ophthalmol Vis Sci.* 1999;40:865–877.

31. Massof RW, Fletcher DC. Evaluation of the NEI Visual Functioning Questionnaire as an interval measure of visual ability in low vision. *Vision Res.* 2001;41:397–413.

32. Velozo CA, Lai JS, Mallinson T, Hauselman E. Maintaining instrument quality while reducing items: application of Rasch analysis to a self-report of visual function. *J Outcome Meas.* 2000;4:667–680.

33. Mangione CM, Phillips RS, Lawrence MG, Seddon JM, Orav EJ, Goldman L. Improved visual function and attenuation of declines in health related quality of life after cataract extraction. *Arch Ophthalmol.* 1994;112:1419–1425.

34. Mangione CM, Orav EJ, Lawrence MG, Phillips RS, Seddon JM, Goldman L. Prediction of visual function after cataract surgery. *Arch Ophthalmol.* 1995;113:1305–1311.

35. Sherwood MB, Garcia-Siekavizza A, Meltzer MI, Hebert A, Burns AF, McGorray S. Glaucoma's impact on quality of life and its relation to clinical indicators: a pilot study. *Ophthalmology.* 1998; 105:561–566.

36. 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.

37. Valbuena M, Bandeen-Roche K, Rubin GS, Munoz B, West SK. Self-reported assessment of visual function in a population-based study: the SEE project. Salisbury Eye Evaluation. *Invest Ophthalmol Vis Sci.* 1999;40:280–288.

38. Kamel HK, Guro-Razuman S, Shareeff M. The activities of daily vision scale: a useful tool to assess fall risk in older adults with vision impairment. *J Am Geriatr Soc.* 2000;48:1474–1477.

39. Rubin GS, Bandeen-Roche K, Huang GH, et al. The association of multiple visual impairments with self-reported visual disability: SEE project. *Invest Ophthalmol Vis Sci.* 2001;42:64–72.

40. Elliott DB, Hurst MA, Weatherill J. Comparing clinical tests of visual function in cataract with the patient's perceived disability. *Eye.* 1990;4:712–717.

41. Rubin GS, Munoz B, Bandeen-Roche K, West SK. Monocular versus binocular visual acuity as measures of vision impairment and predictors of visual disability. *Invest Ophthalmol Vis Sci.* 2000;41: 3327–3334.

42. Linacre JM. *A User's Guide to Winsteps: Rasch-Model Computer Program.* Chicago: Mesa Press; 2002.

43. Wright BD, Masters GN. *Rating Scale Analysis.* Chicago: Mesa Press; 1982.

44. Smith RM. Person fit in the Rasch model. *Educ Psychol Meas.* 1986;46:359–372.

45. Wright BD. Comparing Rasch measurement with factor analysis. *Struct Equat Model.* 1996;3:3–24.

46. Linacre JM. Investigating rating scale category utility. *J Outcome Meas.* 1999;3:103–122.

47. Raju NS, Laffitte LJ, Byrne BM. Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. *J Appl Psychol.* 2002;87:517–529.

48. Monestam E, Wachtmeister L. Impact of cataract surgery on visual acuity and subjective functional outcomes: a population-based study in Sweden. *Eye.* 1999;13:711–719.

49. Laidlaw DA, Harrad RA, Hopper CD, et al. Randomised trial of effectiveness of second eye cataract surgery. *Lancet.* 1998;352: 925–929.

50. Kwapiszeski BR, Gallagher C, Holmes JM. Improved stereoacuity: an indication for unilateral cataract surgery. *J Cataract Refract Surg.* 1996;22:441–445.

51. Javitt JC, Steinberg EP, Sharkey P, et al. Cataract surgery in one eye or both: a billion dollar per year issue. *Ophthalmology.* 1995;102: 1583–1592;discussion 1592–1593.

52. Shuttleworth GN, Luhishi EA, Harrad RA. Do patients with age related maculopathy and cataract benefit from cataract surgery? *Br J Ophthalmol.* 1998;82:611–616.

53. Armbrecht AM, Findlay C, Kaushal S, Aspinall P, Hill AR, Dhillon B. Is cataract surgery justified in patients with age related macular degeneration?—a visual function and quality of life assessment. *Br J Ophthalmol.* 2000;84:1343–1348.

54. Desai P, Reidy A, Minassian DC. Profile of patients presenting for cataract surgery in the UK: national data collection. *Br J Ophthalmol.* 1999;83:893–896.

55. Brenner MH, Curbow B, Javitt JC, Legro MW, Sommer A. Vision change and quality of life in the elderly: response to cataract surgery and treatment of other chronic ocular conditions. *Arch Ophthalmol.* 1993;111:680–685.

56. Laidlaw DAH, Haddad R. Can second eye cataract surgery be justified? *Eye.* 1993;7:680–686.

57. Leinonen J, Laatikainen L. Changes in visual acuity of patients undergoing cataract surgery during the last two decades. *Acta Ophthalmol Scand.* 2002;80:506–511.

58. DeVellis RF. *Scale Development Theory and Applications.* Newbury Park, CA: Sage; 1991.

59. Bland JM, Altman DG. Statistics notes: validating scales and indexes. *BMJ.* 2002;324:606–607.

60. Cortina, JM. What is coefficient alpha?—an examination of theory and applications. *J App Psychol.* 1993;78:98–104.

61. Haegerstrom-Portnoy G, Schneck ME, Lott LA, Brabyn JA. The relation between visual acuity and other spatial vision measures. *Optom Vis Sci.* 2000;77:653–662.

62. McHorney CA. Generic health measurement: past accomplishments and a measurement paradigm for the 21st century. *Ann Intern Med.* 1997;127:743–750.