

Evaluating psychometric properties of a clinical and a self-report blind rehabilitation outcome measure

Judi Babcock-Parziale, PhD;^{1–2*} Patrick E. McKnight, PhD;² Daniel N. Head, EdD¹

¹*Southern Arizona Department of Veterans Affairs Health Care System, Southwestern Blind Rehabilitation Center (3-124), Tucson, AZ;* ²*University of Arizona, Department of Psychology, Tucson, AZ*

Abstract—This study assessed the psychometric properties and evaluated the compatibility of two blind and low-vision rehabilitation outcome instruments, the VA-13 and the Functional Assessment of Self-Reliance on Tasks (FAST). Legally blind veterans ($N = 190$) from a Department of Veterans Affairs inpatient blind rehabilitation center completed the VA-13 (a retrospective pretest and posttest) at 6 weeks postdischarge. Clinicians rated veterans on the FAST at admission and discharge. The psychometric properties of the two instruments and their compatibility were evaluated with the use of a Rasch model analysis. The two instruments functioned consistently as screens and showed a ceiling effect at posttest; however, the VA-13 showed poor sensitivity to change. In contrast, the FAST showed more reliable change, but a few items changed in unexpected ways. Our conclusions show that the two instruments are currently not compatible for calibration; however, this can be improved with proper attention to scaling inadequacies, test administration times, and content coverage.

Key words: aged, blind rehabilitation, clinician ratings, low vision, measurement, psychometrics, Rasch model, rehabilitation, self-report ratings, treatment outcomes.

INTRODUCTION

This study assessed the psychometric properties of two blind rehabilitation outcome measures and evaluated the compatibility of the measures using the same population of inpatient blind rehabilitation patients enrolled at a Department of Veterans Affairs (VA) blind rehabilitation center (BRC). The two measures are a clinical rating

instrument—the Functional Assessment of Self-Reliance on Tasks (FAST)—and a self-report rating instrument—the VA-13. To determine if the two are compatible and produce similar results, we needed to estimate the psychometric properties of both measures for the same group of patients. Outcome instruments from the same field are often assumed, rather than empirically determined, to be compatible when they contain similar questions about the same content. This is certainly the case in the field of blind and low-vision rehabilitation in which numerous outcome instruments have emerged in the last decade [1–2].

Low-vision and blind rehabilitation outcomes need to be compared across studies to improve clinical practice and advance this field of scientific inquiry. Linking clinician

Abbreviations: BRC = blind rehabilitation center, CI = confidence interval, CTI = classical test theory, FAST = Functional Assessment of Self-Reliance on Tasks, IRT = Item Response Theory, SD = standard deviation, SWBRC = Southwestern Blind Rehabilitation Center, VA = Department of Veterans Affairs.

This material was based on work supported by the Department of Veterans Affairs Rehabilitation Research and Development Service to Dr. Babcock-Parziale, grant C-2710I.

*Address all correspondence to Judi Babcock-Parziale, PhD; Southern AZ VA Health Care System, Southwestern Blind Rehabilitation Center (3-124), 3601 S. 6th Avenue, Tucson, AZ 85723; 520-792-1450, ext. 5698; fax: 520-629-4995. Email: judith.babcock@med.va.gov

DOI: 10.1682/JRRD.2004.07.0080.

ratings with patients' self-ratings may also provide a more comprehensive assessment of the functional improvement attained by patients. The FAST ratings, for example, denote the efficacy of the blind rehabilitation training, and efficacy should be related to the VA-13 effectiveness ratings. Additionally, establishing baseline differences between clinician and self-ratings and empirically determining the sources of these differences are important.

While most researchers may consider that different respondents provide qualitatively and quantitatively different responses, adequate evidence exists to support the agreement between clinician and patient ratings [3–5]. An individual's functional ability consists of signs, markers, and behaviors, and we would not expect the two respondents' ratings to be totally different or unrelated. If, however, the two respondents provide completely different ratings for the same construct, then those differences may be explained by differences in the instruments (i.e., the way the questions are worded or responses are coded). On the other hand, if the instruments share sufficient commonality, they could be calibrated into a common scalar [6]. Calibration permits researchers to substitute one measure for another (e.g., ratings on one instrument could be converted to a score on the other instrument), thereby making programs comparable even when different outcome measures are used. Therefore, our secondary aim was to determine if the two measures meet the criteria for calibration.

INSTRUMENTS

The VA-13

The VA-13, formerly referred to as the Blind Rehabilitation Service Follow-up Outcomes Survey, is a 13-item self-report measure of the frequency of, independence in, and satisfaction with performing specific tasks [2,7].* The VA-13, developed by researchers at the VA Atlanta Rehabilitation and Research Center for the VA Blind Rehabilitation Service, has been the standard used for the past 6 years by the service to account for the effectiveness of 10 inpatient BRCs. The VA-13 consists of ques-

* A subset of VA-13 items addresses frequency of tasks and satisfaction with performing specific tasks, but these items were not relevant to this study, which focused on the independence scale.

tions that sample self-reported behaviors and perceptions associated with tasks that are linked to each of the four major domains of vision rehabilitation skills.† The instrument is administered by research staff to veterans as a posttreatment telephone survey 4 to 6 weeks after they are discharged from a BRC. At the time of the posttreatment interview, the veteran is asked to rate on a scale of 1 to 3 his or her current ability to complete each of the 13 tasks, with a "1" indicating the task could be completed with a great deal of assistance, a "2" indicating a little assistance, or a "3" indicating no assistance (independently). Following the report of current ability (posttest), the veteran is then asked to rate his or her ability to perform the same task before enrolling in blind rehabilitation (a retrospective pretest). A change score is computed with the raw scores (the retrospective pretest score is subtracted from the posttest score).

Functional Assessment of Self-Reliance on Tasks

The second outcome instrument is the FAST, an 11-item clinical measure developed by researchers at the VA's Southwestern BRC (SWBRC) in Tucson, Arizona. Colenbrander notes, when we talk about "disability," we are actually discussing "ability" [8]. Massof also discusses this point in his study [9]. The FAST was developed to assess change in functional ability, and researchers and blind rehabilitation instructors determined its content theoretically and empirically [10]. The 11 items were written to express content-relevant information at the goal level (i.e., general activity categories where daily activities or tasks may be grouped) [11]. A veteran is rated on the FAST through consensus scoring by a team of clinicians. SWBRC has four transdisciplinary teams that are composed of instructors representing the four training disciplines. The teams meet daily to discuss a veteran's training, and at admission and discharge, teams use a consensus process to score the FAST items. The consensus process establishes the interrater

† The VA blind rehabilitation training program consists of four disciplines: Low Vision, Orientation and Mobility, Living Skills, and Manual Skills. The Manual Skills training, unique to the VA, is designed to assess and enhance skills in all aspects of sensory awareness with an emphasis on adaptive and safety techniques. Instructional areas include leatherwork, copper tooling, home mechanics, small engine repair, and woodworking. It is not considered vocational training although some veterans have developed a vocation or hobby from it.

reliability within the team. Interrater reliability across teams is monitored quarterly with the selection of a group of veterans from each team (i.e., matched on visual acuity range, visual field, age, and medical comorbidities). If any one of the four teams' ratings is significantly different on any of the FAST items for the sample of veterans, then the rating discrepancies are identified and reported to the team leader. If necessary, team leaders review the FAST scoring procedures with team members and the consensus process.

The FAST was designed to serve two functions: first, the instrument serves as a clinical screen, providing clinicians with the assessment information required to develop both treatment and discharge plans. Second, it measures the change in functional ability, which is defined as a blind or severely visually impaired veteran's ability to perform the instrumental activities of daily living [12] that are required to promote personal independence and fulfill traditional social roles. To rate the goal, Food Preparation, for instance, clinicians consider a veteran's ability to perform typical tasks including organizing and planning a menu; preparing cold foods; preparing hot foods using either the stove top, oven, or microwave; serving the meal; adaptive eating, and cleaning the kitchen. Veterans were rated on the 11 FAST goals with the use of a 10-point scale, which ranges from "1"—unable to perform the tasks or is a danger to self when attempting to perform tasks to "10"—excellent skills demonstrated with complete independence in all aspects of the item. Four levels of functioning categorize the clinically judged criteria: Dependence (ratings 1–3), Limited Dependence (ratings 4–5), Limited Independence (ratings 6–8), and Independence (ratings 9–10). Ratings are based on the use of low-vision devices and other adaptive aids while one performs these tasks, and in most cases, the use of the adaptive aids are reflected only in the posttest ratings. Clinicians have used the FAST for the past 6 years to evaluate SWBRC program outcomes.

Both instruments were field tested before this study, and those results indicate that the VA-13 measures functional independence and the FAST measures functional ability [2,10]. The VA-13 tasks were identified by an expert panel that included clinicians, and theoretically, these tasks should correspond with the broader FAST goals, since both instruments are based on the standard four-domain VA blind rehabilitation training curriculum. For example, the VA-13 asks veterans to rate their ability to measure using common kitchen measuring devices.

This task was hypothesized to be part of the broader FAST goal, Food Preparation. While the two instruments appeared to measure similar constructs, little was known about the compatibility of the two measures. Furthermore, it was important to determine if both instruments are suitable measures of change that can be used by the VA Blind Rehabilitation Service to account for blind rehabilitation program outcomes at both the clinical and patient level. The present article illustrates the process we used to evaluate the psychometric properties for each instrument and to determine whether the two instruments were compatible and qualified for calibration.

Development Limitations

Assessment of the validity and reliability of the VA-13 and the FAST must be viewed through the "real-world" lenses of applied measurement. Given the pressing need of developing outcome measures for the field of blind rehabilitation, the two instruments moved quickly from the development stage to the measurement practice stage. We need to note an important distinction between measurement development and measurement practice. The traditional measurement development process provides the researcher with the opportunity to develop and refine the instrument using a theory-driven process and to pay vigilant attention to the process of measurement. In measurement practice, the instrument is used shortly after its development and may be the only data regularly collected, so finding evidence of problems can be quite difficult [13]. Both of the measures examined in this study fall into the category of measurement practice. While sufficient attention was given to measurement processes, the development process had limitations imposed by the urgent need for national outcomes data and the restrictions imposed by clinical practice. The comparison and contrast of the psychometric results presented in this article demonstrate the methodology used to identify areas in each instrument that require additional development.

METHODS

Subjects

The sample included veterans who attended the SWBRC inpatient blind rehabilitation program between December 2000 and July 2002. The VA-13 and the FAST databases composed of the same sample of veterans were merged, yielding 190 cases matched for secondary analyses.

The median age of the sample was 77 years (range 42–96). The sample consisted of legally blind veterans who were married (60%), male (93%), and white (85%) and who had sought rehabilitation services for their visual problems. Over half (57%) lived with their spouse, and 14 percent had not received a high school diploma.

Veterans admitted to the program were legally blind (i.e., acuity of 20/200 or worse) as determined with the use of the Bailey Lovie Distance Chart [14]. The primary diagnoses included age-related macular degeneration (66%), glaucoma (10%), diabetic retinopathy (7%), retinitis pigmentosa (4%), and other (13%). In terms of visual field loss, 55 percent had a central field loss, 10 percent peripheral loss, 14 percent both central and peripheral, and 21 percent no field loss.

The entering mean score for the binocular log-visual acuity of the better eye was 1.3, and the mean of the best-corrected log-visual acuity at discharge was 1.2. Thus, as a result of new refraction, patients' acuity in the better eye improved by one line on the distance chart.

Psychometric Evaluation of Measures

The two scales were analyzed with a traditional psychometric approach, classical test theory (CTT), and then subsequently with a Rasch model analysis. Classical test scales are made up of the true score plus some amount of error [15]. The true score is the expected value of the observed score (i.e., the average score if the test could be administered to the same person multiple times) [9]. Sentiments have grown in educational and psychological testing to abandon the CTT approach in favor of the more modern latent response models (i.e., Item Response Theory [IRT] and Rasch) [9,15]. "Where CTT begins with a test score, IRT begins with an explicit mathematical model that describes the relationship between responses to the instrument's items and the trait of the person to be measured" [9, p. 531]. The authors acknowledge the limitations of CTT; however, the descriptive statistics, Cronbach's α and intra-class correlation coefficients for the VA-13 and FAST are included to allow readers to compare these data with previously published data [2,10] (results can be viewed in **Appendix Tables 1 and 2**, available online only at www.vard.org/jour/jourindx.html).

Rasch Methods

To evaluate the two instruments, we used Rasch model analysis, also referred to as a one-parameter IRT model [15]. The Rasch model provides estimates of both

items and persons independently. Specifically, an individual's ability is estimated independently from the item difficulty estimates [16]. In contrast, CTT estimates item difficulty and person abilities together, thereby confounding the parameter estimates. Rasch models have been successfully applied to visual function assessments (as well as a variety of other rehabilitation and medical outcome areas), and studies demonstrate that valid interval scales for visual ability can be constructed from rating scale responses to items by visually impaired people [17–19].

The Rasch rating scale was used to transform the raw scores into equal interval logits or the log odds of the probability of an observation actually occurring [$\log(P/1 - P)$] [16,20]. The Rasch model is a conjoint (additive) probability model that estimates person and item difficulties using the maximum likelihood estimation for each element specified in the model. The elements specified for these analyses were patient ability and test item difficulty. The model is based on the logical assumption that patients with high levels of functional ability should have an increased probability (relative to patients with low levels) of getting a higher score on an item.

We examined the precision of the two instruments by comparing the variance of the test items with the variance of the sample, which is referred to as separation reliability. Separation reliability informs us how many levels of functional ability can be reliably detected in the sample given the FAST and VA-13 test items. Several strategies were also used for assessing construct validity. First, we determined how well the FAST and VA-13 measured what they purport to measure (i.e., functional ability and independence). If responses describe functional ability and independence meaningfully, then higher-functioning patients would manifest higher response category usage, while lower-functioning patients would demonstrate lower response category usage. Second, we examined the fit indices and items calibrations for each time point. We assessed the mean square fit statistics, which indicate the ratio of expected to observed variance, to determine how well an item fits the underlying construct. Evidence of construct validity, for example, included how well each FAST and VA-13 test item reflects coherence between item ratings and the overall state of functional ability and whether each item provides independent information about functional ability. Finally, we assessed the construct validity by evaluating the stability of the FAST and VA-13 over time. That is, to illustrate stability of the items over

time, we examined item calibrations for each time point by plotting them from pretest to posttest.

Finally, the Rasch model requires unidimensionality [20], and although unidimensionality has no formal test, a single factor consistently emerged from an exploratory factor analysis using both principal factor (i.e., varimax rotation) and principal components analyses for both instruments [16]. Based upon the factor analysis results, we found no overwhelming evidence to lead us to reject unidimensionality in the VA-13 and FAST.

RESULTS

Rasch Analysis of VA-13

The Rasch person-item map is a common ruler that graphically depicts patients' self-reported ability to perform selected daily tasks (i.e., person measure) in parallel with the visual ability needed to perform the tasks (i.e., item measure). We examined the person-item maps to compare the range and position of the item measures with the person measures and to determine the amount of logit change. The VA-13 person-item maps created for both the retrospective pretest and posttest scores can be viewed in the **Appendix Figures 1 and 2** (available online only at www.vard.org/jour/jourindx.html). The distribution of person scores for the retrospective pretest has noteworthy problems (negatively skewed and slightly kurtotic). The mean of the persons is 1 logit greater than the mean of the items, which indicates that the retrospective pretest fails to capture patients with higher levels of visual ability (i.e., the retrospective pretest has a ceiling effect). As a result, the person-item map indicates that about 41 percent of the patients report functional independence on the 13 tasks before rehabilitation.

At posttest, the dispersion of items stays about the same (3 logits), and the person measures have shifted upward by 1 logit. Examining logit change is important because it informs us about how sensitive the instrument is to change. A change in logit(s) is a metric change and should not be confused with an estimate of clinical significance. In this case, the logit change is similar to an effect size (Cohen's d), and the change for the VA-13 is $d = 0.63$. In other fields, this moderate effect signifies above-average change. However, previous outcome studies in our field have documented closer to 2, and in some cases, almost 4 logits of change for similar tasks using the self-report method [5,8].

VA-13 Item Measure Characteristics

Before blind rehabilitation, a number of tasks such as reading a newspaper or crossing a street with a traffic light were especially difficult, if not impossible, for a legally blind veteran. Therefore, we would expect the item difficulties to reflect the same order of difficulty that is observed in clinical practice at admission or in pretest self-reports. The distribution of item difficulties represented in the retrospective pretest person-item maps is inconsistent with previous clinical observations and previous research. The most difficult activity reported in the VA-13 retrospective pretest is Item 3—Read Mail, which is about 1 logit more difficult than Item 8—Read a Magazine or Newspaper. This order of difficulty is similar in the posttest scores. Sustained reading of a newspaper or magazine, however, is more difficult for patients than reading mail whether it is measured at admission or discharge. Blind rehabilitation instructors report that the small print and the sustained reading of newspapers or magazines makes reading more difficult than reading mail. Furthermore, the map of persons and items reported by Stelmack et al. demonstrates that reading a newspaper or magazine is a more difficult than reading mail [19, p. 238]. While these two items are disordered, the remaining VA-13 items are ordered according to their expected difficulty. We also observed that the dispersion of items for both pre- and post tests is limited (i.e., the item difficulties are all about the same), which indicates that the VA-13 resembles a screening instrument [21].*

An additional method used to examine the item measures was a pretest/posttest graphical analysis of the VA-13 (i.e., retrospective pretest and posttest). **Figure 1** illustrates that all items fall on the identity line (within the confidence interval [CI] represented by the dotted lines), indicating that these items do not change over time as would be expected. However, two VA-13 items fall on the border of the CI. Item 3, Read Mail, becomes easier at posttest and shows the most change. Item 4, Watch TV, actually becomes more difficult, which is counterintuitive. This finding may be because veterans are watching less television and therefore are rating it as more difficult.

*A *screen* refers to the ability of the instrument to order items. If an instrument cannot order respondents but does a reasonable job at discriminating between two meaningful levels of responses, it is considered to be a screen.

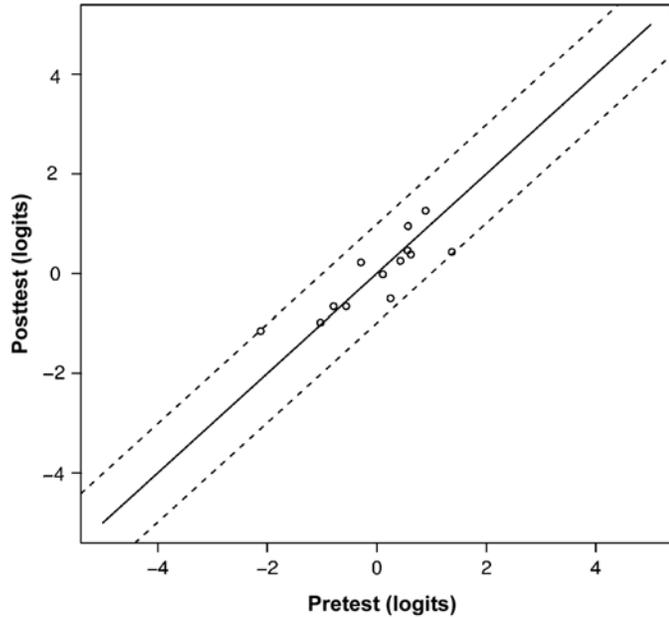


Figure 1.
VA-13 item measures.

The fit statistics (infit and outfit) are indices of measurement accuracy. Bond and Fox suggest that reasonable item mean square ranges for infit and outfit statistics are 0.6 to 1.4 for rating scales (VA-13) and 0.5 to 1.7 for clinical observation (FAST) [16, p. 179]. Overall, the fit statistics for the VA-13 fell within the range that fits the Rasch model for both retrospective pretest and posttest treatment administration. The infit mean squares for the retrospective pretest range were from 0.76 to 1.43 and the posttest range was from 0.87 to 1.29. More details about the fit statistics for each item by time are included in **Appendix Table 3** (available online only at www.vard.org/jour/jourindx.html).

VA-13 Person Measure Characteristics

The scatter plot of retrospective pretest and posttest person measures is displayed in **Figure 2**. The plot illustrates that almost half of the patients were above zero (y-axis), and this is the group of patients who reported the ability to complete the 13 tasks before rehabilitation. The strong clustering of the person measures along the identity line indicates that the instrument provides evidence of patient change due to rehabilitation services for the 59 percent of patients who were not able to complete the tasks before rehabilitation. Finally, the clustered scores in the top of the plot indicate the ceiling effect in the rating scale.

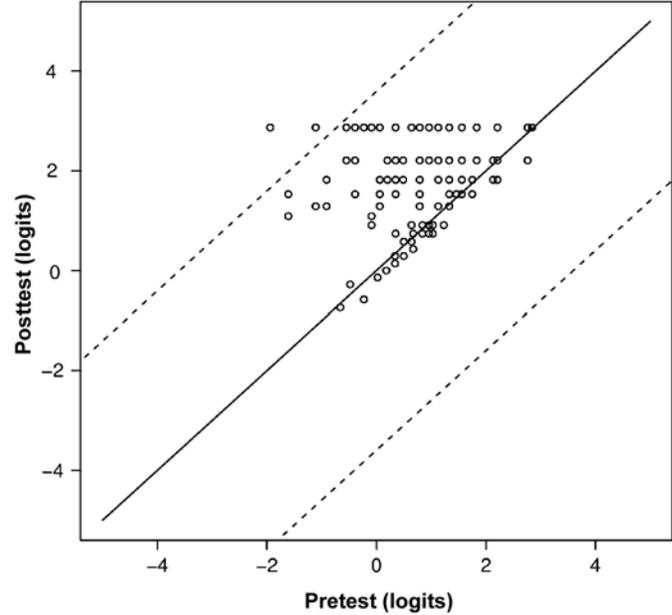


Figure 2.
VA-13 person measures.

Reliability of Items and Persons

The Cronbach's α 's for the VA-13 items of the retrospective pretest and posttest data were 0.81 and 0.76, respectively. These estimates indicate coherence among the items for each administration. The person reliability estimates for the two time administrations are 0.71 and 0.27. Although the VA-13 has good-fit indices, the 0.27 person reliability (Real root-mean square error) indicates poor person reliability at posttest (see **Appendix Table 3** for fit indices, available online only at www.vard.org/jour/jourindx.html). The low person estimate at posttest also indicates that 73 percent of the variance is error (e.g., random noise). The contradiction between the CTT and Rasch reliabilities occurs because Cronbach's α includes all the people in the ceiling as being measured perfectly, thereby inflating the reliability. Rasch deletes the scores in the ceiling that are invariant, which deflates the reliability. A good model fit has occurred because there is little to measure and the error (or noise) is symmetric; hence, the actual response corresponds perfectly to the expected within the limits of error. As a result, the poor sensitivity to detect change enables the VA-13 to appear to have good reliability when Cronbach's α is used.

The person-separation reliability is the adjusted person standard deviation (SD) divided by the average measurement error. This ratio is useful for determining how many

distinct strata or groups in which the person measure distribution can be divided. The person separation index for the retrospective pretest is 1.57 and the posttest it is 0.60. Both tests are below the 2.0 minimum criterion for a screen (i.e., an instrument used to discriminate between at least two groups), and these estimates indicate that the instrument is not discriminating between patients with different levels of functional independence.

VA-13 Scale Properties

The VA-13 scale properties were evaluated for both administrations of the scale. An extreme distribution of scores indicates that most patients responded to the items using primarily one side of the rating scale (i.e., “3” or completely independent). The 3-point scale functions as a dichotomous scale with most of the patient ratings occurring at steps 1 and 3 for both the retrospective pretest and posttest. These findings are consistent with our previous findings of rating scale usage on an independent sample of the VA-13.*

Rasch Analysis of the FAST

The Rasch person-item maps for FAST pre- and posttest administrations are available in **Appendix Figures 3** and **4** (available online only at www.var.org/jour/jourindx.html). The distribution of item difficulties is consistent with both clinical reports and the item ordering identified by Stelmack et. al [19, p. 238]. Qualitatively speaking, the person measures are normally distributed across the items at pretest. In addition, at pretest, the mean of the patients is about 1 logit lower than the mean of the items, with an SD of about 2.5 logits. The difference indicates that the majority of the patients (about 94%) did not have the functional ability to complete the 11 FAST goals before rehabilitation. At posttest, the distribution of patient scores had a strong positive skew and the patient mean shifted substantially above the item difficulty level, indicating that patients performed at almost 2.5 logits above the item mean. This logit change corresponds with a very large effect size, $d = 1.8$. The large logit shift (and effect size) indicates that following blind rehabilitation, 77 percent of the patients increased their functional ability to the

level required to complete the 11 FAST goals with above average functional ability (i.e., scores > 7 on the FAST).

FAST Item Measure Characteristics

Both infit and outfit statistics for FAST item measures were consistent with the assumptions of the Rasch measurement model for clinical ratings [16] (0.50–1.7), except for the pretest Reading item (1.82). The infit mean squares range between 0.74 and 1.22 at pretest and between 0.82 and 1.35 at posttest (**Appendix Table 4**, available online only at www.var.org/jour/jourindx.html).

The FAST plot of pre- and post-items provided in **Figure 3** illustrates that three of the items fall outside of the dotted line or the CI range (+2 standard error). The fit statistics for the item measures indicate that all FAST items fit well, except pretest Item 8, Reading. We hypothesized that the “noise” in the Reading item at pretest was most likely due to two unique characteristics of the patients. We attempted to predict the error in the Reading item with the use of visual acuity and central field loss, but neither predictor helped to explain the poor fit. Our current, unsubstantiated explanation is that the noise in the reading item represents an amalgam of patient preferences for reading, premorbid reading ability, and other idiosyncratic factors that cannot be captured in our current data.

The FAST Item 2, Home Maintenance, and Item 3, Fine Motor/Dexterity, show a small pre- to postchange. These goals represent training areas that are not easily improved, especially given the physical limitations of our geriatric patients. These items appear to become more difficult in relation to the other items; however, the majority of patients who choose training in these two areas do show some improvement.

The plot of items in **Figure 3** also reveals the limited dispersion of the FAST items, which indicates this instrument also performs as a screen. Although the item difficulty levels are restricted in a screening instrument, the 10-point rating scale is used consistently across the instrument and this accounts for the variability observed for the scale.

FAST Person Measure Characteristics

The FAST pretest and posttest person measures are plotted in **Figure 4**. The plot indicates that patients increase in functional ability over time. Patients’ lower and upper ranges of ability are accounted for, but some evidence of a ceiling effect is shown. The FAST person-item

*Babcock-Parziale J, Head DN, McKnight P, Massof R. Calibration of clinical and self-report measures in blind rehabilitation. Invited poster presented at the VA Rehabilitation Research and Development 3rd National Meeting; 2002 February; Arlington, Virginia.

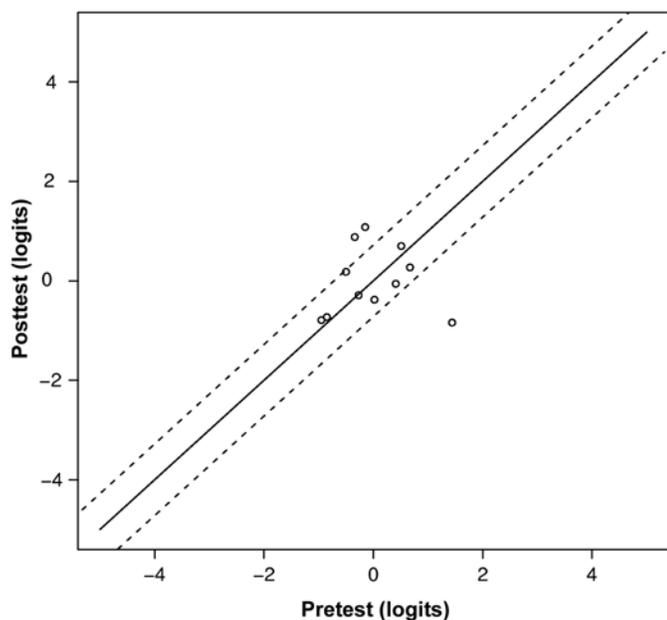


Figure 3. Functional Assessment of Self-Reliance on Tasks (FAST) item measures.

characteristics reported in this study replicate findings from a previous analysis of over 500 FAST cases.*

Reliability of Items and Persons

The reliability estimates of the item measures for the FAST pretest and posttest data are 0.97 and 0.95, respectively. These high and consistent reliability scores indicate that the FAST has good measurement precision. The person reliability estimates for the two time administrations are 0.90 and 0.85. Thus, the FAST is discriminating between persons and accounts for 85 percent of the variance in the distribution.

The separation index for the pretest is 2.95 and the posttest is 2.37. Both administrations are above the 2.0 minimum criterion for a screen; hence, the measure is able to discriminate between those patients who are and those who are not able to complete the functional goals.

FAST Scale Properties

All categories of the 10-point scale are endorsed; however, the lower steps (categories 1–4) are more likely to be

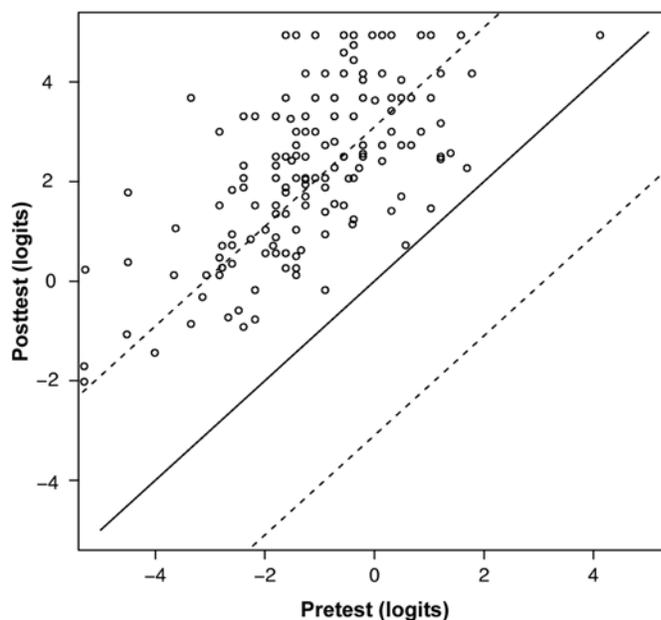


Figure 4. Functional Assessment of Self-Reliance on Tasks (FAST) person measures.

endorsed at pretest, while the higher categories (7–10) are endorsed more often at posttest. While the item cluster indicates a screen, the rating scale provides enough variability to differentiate the items sufficiently to produce reasonable judgments about successful attainment of the functional goals associated with blind rehabilitation.

DISCUSSION

We conducted an analysis on the raw scores for both instruments using CTT, but we did not form any strong conclusions from these results because we needed more information about the scaling properties. To this end, we used a Rasch model analysis to transform the VA-13 and FAST scores into interval-like scales and to provide other diagnostic information that is somewhat available in CTT but provided by default with the Rasch analysis. Based upon the Rasch model results, we offer the following conclusions for the two instruments.

First, the VA-13 retrospective pretest has a strong ceiling effect in this current sample as well as in previous and subsequent samples. The ceiling effect indicates a restriction of range problem; therefore, the VA-13 might be a poor outcome measure due to the low variability

* Babcock-Parziale J, Head DN, McKnight P, Massof R. Validation of the Functional Assessment of Self-reliance on Tasks. Invited paper presented at the American Academy of Optometry; 2002 December; San Diego, California.

from pre- to postrehabilitation. Results show that 41 percent of patients who completed the VA-13 retrospective pretest were able to complete the tasks before admission, which indicates that almost half the patients may not require blind rehabilitation services. The FAST pretest scores for the same group of patients revealed that 6 percent of the patients were able to adequately complete all the goals before admission. Additional research is required to determine whether the traditional pretest is more sensitive than the retrospective pretest when the VA-13 is used to measure patients' functional ability before blind rehabilitation training.

Second, the VA-13 has poor sensitivity to change—a problem that may be due to the aforementioned ceiling effect. The low person-reliability estimate at posttest (0.27) and the discrepant Cronbach's α (0.91) indicate the items are poorly suited for the sample and, consequently, for the population. An unreliable estimate at both pretest and posttest also makes the estimated logit change unreliable [22]. The VA-13 exhibited a mean change of 1.1 logits (SD = 1.2), and the change scores illustrated in **Appendix Figure 5** (available online only at www.vard.org/jour/jourindx.html) might be best characterized as a Poisson distribution rather than a normal distribution. Hence, patients tended to change very little from pre- to postrehabilitation. The FAST scores for the same group of patients illustrated in **Appendix Figure 6** (available online only at www.vard.org/jour/jourindx.html) were more normally distributed and showed a greater range and variability of patient change. The FAST scores reflect an average mean change of 3.6 logits (SD = 1.5). The difference between the VA-13 and the FAST change scores may be due to the respondent; however, no evidence exists to suggest that the entire difference in the amount of change observed is due to respondent bias (self-ratings vs. clinician ratings). As previously noted, studies in this field that used self-report instruments have documented anywhere from a 1- to 4-logit change—a range consistent with the change scores observed in the present study for both clinician and self-report.

In addition to the ceiling effect that occurs for both the retrospective pretest and posttest scores, the restricted use of the rating scale may account for the lack of variability and the minimal difference in observed change in the VA-13. Hence, our third conclusion is that the 3-point scale does not appear to capture the full variability in patient functioning that it was intended to capture. One explanation is that patients may be selecting the midpoint (rating a "2" on the 3-point scale) as a point of indecision rather than using it to

denote a mid-level of independent functioning. The highest value (i.e., the rating value "3") tends to be endorsed more often than the other two rating options, indicating either a bias in the response category usage or a restriction of scaling options suitable for the patient population.

While the Rasch results for the FAST indicate that the person estimates change reliably from pretest to posttest, the instrument does have several troublesome properties. First, the FAST provides the expected shift from unable to able to complete the tasks over the course of the blind rehabilitation program, but just as we observed in the VA-13, it too has a ceiling effect. The FAST fails to account for much variability in the posttest administration where veterans tend to reach a similar level of functioning. Despite the ceiling effect, the Rasch model results indicate the FAST offers a separation index greater than 2.0 and, therefore, may reliably differentiate between successful and unsuccessful rehabilitation outcomes. Since the ceiling effect is observed in the posttest administrations of both the FAST and the VA-13, we suspect the ceiling effect may be attributed to the nature of a "screen" (i.e., the restricted range results from similar item difficulty levels).

A second troubling FAST finding pertains to Items 2 and 3; both items appear to be more difficult after the rehabilitation process. Although item difficulty levels should remain the same, the finding may be logically defended given that Home Maintenance and Fine Motor Skills are goals that are not as easily improved as is Reading in this geriatric population. This may also be the case for VA-13 Item 4, Watching TV, which also becomes more difficult, which may be why these activities are not engaged as often and are not a priority for the veterans. At any rate, the two FAST items must be revised so that their difficulty levels remain consistent over time.

Finally, FAST Item 8, Reading, and VA-13 Item 3, Read Mail, both became easier at posttest and demonstrate the most change. Reading is a task that is most amenable to change when patients receive the appropriate magnification aids and are properly trained to use them. The large shift in Reading item parameters can be attributed to the veterans not having these aids at pretest, when most legally blind veterans are unable to read. At posttest, veterans were trained to appropriately use the reading aids, and the reading tasks became much easier. The introduction of the aids may in effect introduce a change in the context of the reporting that would explain why the item difficulty changed so dramatically. Nevertheless, retaining the Reading item is important because

improving access to reading is the primary goal selected by the majority of veterans with low vision. We will therefore revise the FAST Reading item to account for the change in the context of reporting and to improve the Rasch model fit of this item.

Earlier we hypothesized that the VA-13 and the FAST might be compatible and suitable for calibration. We anticipated that we could evaluate the continuity and effectiveness of the inpatient blind rehabilitation program at both the clinician and patient level. The Rasch model analyses, however, directed our attention to the differences between the instruments. The differences include the level of measurement, time of administration, rater, number of items, and the rating scale (**Table**).

We were not able to determine with these data whether the lack of agreement between the VA-13 retrospective pretest and the FAST pretest can be attributed to the time of administration. To determine whether the retrospective scores can be substituted for pretest scores and used to compute a change score, one must compare the VA-13 retrospective pretest with a pretest administered before admission. Additionally, the Rasch model analysis highlighted the effect that the differences might have on rating scale usage and, consequently, on the psychometric properties. Clearly, the rating of FAST goals by instructors (based on a consensus rating by experts on a team) produced more variability than the patients' self-ratings of a single task. At any rate, the differences in rating scale usage and the lack of variability produced by the VA-13's 3-point scale were the main issues that made the calibration of the instruments problematic at this time. We anticipate that revising both instruments will improve future calibration efforts and provide a better estimate of functional improvement.

Finally, we conducted this research study based on the premise that the respondent facet would not influence the change scores. In other words, we had little reason to suspect that the two instruments would differ simply because of the respondent. We continue to hold this view given the fact that the instruments differed for many reasons—none of which were manipulated or tested in the

current study. Although these differences exist for the between-measure calibration, the estimates derived for each instrument are a sound starting point for researchers to better understand the relationship between the values obtained for self-report and clinical ratings.

CONCLUSIONS

The evaluation of the psychometric properties of each instrument was a valuable procedure, and it is an essential part of any measurement process because it provides insight into the characteristics of the instruments and helps further development. The complete battery of analyses indicates the current and analyzed version of the VA-13 showed poor sensitivity and probably underestimated patient change. The FAST is a suitable instrument to screen patients for rehabilitation potential, and it may also be useful for assessing program performance (e.g., quality control or interprogram comparison) at the nine other VA BRCs.

The VA-13 and the FAST can both be improved with proper attention to scaling inadequacies, test administration times, and content coverage. While all methods of data collection may present inherent differences between instruments, assessing the extent that these differences may contribute to variance in item difficulty estimates and patient ability estimates is important. Even more important is a better understanding of the validity of the items; clearly, these results show that the two instruments are not compatible in their current form. Future research to determine the effectiveness of blind rehabilitation training should focus on validating the extent of disability and how well the instruments capture disability using both self-report and clinical ratings.

ACKNOWLEDGMENTS

We wish to acknowledge and thank the veterans who participated in this study; without their participation, scientific and clinical knowledge would not advance. We also

Table.
Differences between measures.

Measure	Item Level	Pretest	Posttest	Respondent	No. of Items	Rating Scale (Points)
FAST	Goal	Admission	Discharge	Clinician rating	10	10
VA-13	Task	Retrospective pretest	Postdischarge	Patient self-report	13	3

wish to recognize the staff at SWBRC for their years of dedicated service to visually impaired veterans and their participation in developing the clinical instrument, the Functional Assessment of Self-reliance on Tasks. We also want to thank Dr. William De l'Aune and Dr. Mike Williams of the Atlanta VA Rehabilitation and Research Development Center for contributing the VA-13 data for those veterans who attend the SWBRC. We note a special thanks to Dr. Robert W. Massof, our project consultant, of the Lions Vision Research Center, Wilmer Ophthalmologic Institute, The Johns Hopkins University School of Medicine. We appreciate his collaboration and his invaluable expertise in low-vision measurement. His support and input helped to advance our theory of measurement for blind rehabilitation outcomes. Finally, we thank the anonymous reviewer who provided a thorough, thoughtful, and helpful critique of our Rasch analysis interpretation.

REFERENCES

1. Stelmack J. Quality of life of low-vision patients and outcomes of low-vision rehabilitation. *Optom Vis Sci.* 78(5): 335–42.
2. De l'Aune W, Williams M, Watson GR, Schuckers P, Ventimiglia G. Clinical application of a self-report, functional independence outcomes measure in the DVA's Blind Rehabilitation Service. *J Vis Impair Blind.* 2004;98(4):281–91.
3. McCabe P, Nason F, Demers Turco P, Friedman D, Seddon JM. Evaluating the effectiveness of a vision rehabilitation intervention using an objective and subjective measure of functional performance. *Ophthalmic Epidemiol.* 2000;7(4):259–70.
4. Agostinho Jn F. Subjective rating of health among the elderly. *Act Adaptation Aging.* 1985;6(4):53–62.
5. Babcock-Parziale J, Cunningham V. Understanding change in self-perceived functional status in veterans following low-vision rehabilitation. Invited paper presented at the American Academy of Optometry; 2003 December; Dallas, TX.
6. Sechrest L, McKnight P, McKnight K. Calibration of measures for psychotherapy outcome studies. *Am Psychol.* 1996;51(10):1065–71.
7. De l'Aune WR, Williams MD, Welsh RL. Outcome assessment of the rehabilitation of the visually impaired. *J Rehabil Res Dev.* 1999;36(4):273–93.
8. Colenbrander A. The visual system. In: Cocchiarella L, Andersson BJ, editors. *Guides to the evaluation of permanent impairment.* 5th ed. Chicago (IL): American Medical Association; 2000. p. 277–304.
9. Massof RW. The measurement of vision disability. *Optom Vis Sci.* 2002;79(8):516–52.
10. Head DN, Babcock JL, Goodrich GL, Boyless JA. A geriatric assessment of functional status in vision rehabilitation. *J Vis Impair Blind.* 2000;94(6):357–71.
11. Massof RW. A systems model for low vision rehabilitation. I. Basic concepts. *Optom Vis Sci.* 1995;72(10):725–36.
12. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist.* 1969;9(3):179–86.
13. Wilson MR. *Constructing measures: An item response modeling approach.* Mahwah (NJ): Lawrence Erlbaum; 2005. p. 172.
14. Bailey IL, Lovie JE. New design principles for visual acuity letter charts. *Am J Optom Physiol Opt.* 1976;53(11): 740–45.
15. Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of Item Response Theory.* Newbury Park (CA): Sage; 1991. p. 2–5.
16. Bond TG, Fox CM. *Applying the Rasch model.* Mahwah (NJ): Lawrence Erlbaum Associates; 2001. p. 188–96.
17. 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(3):397–413.
18. Stelmack J, Szlyk JP, Stelmack T, Ardickas Z, Massof RW. Sensitivity of the VA LV VFQ-48 to change after low vision rehabilitation. Presented at the Association for Research in Vision and Ophthalmology [program 3816]; 2002; Ft. Lauderdale, FL.
19. Stelmack J, Szlyk JP, Stelmack T, Babcock-Parziale J, Demers-Turco P, Williams TR, Massof RW. Use of Rasch person-item map in exploratory data analysis: A clinical perspective. *J Rehabil Res Dev.* 2004;41(2):233–41.
20. Rasch G. *Probabilistic models for some intelligence and attainment tests.* Copenhagen (Denmark): Danmarks Paedagogiske Institute; 1960.
21. Lord FM. *Applications of Item Response Theory to practical testing problems.* Educational Testing Service. Hillsdale (NJ): Lawrence Erlbaum; 1980.
22. Rogosa DR, Willett JB. Demonstrating the reliability of the difference score in the measurement of change. *J Educ Meas.* 1983;20(4):335–43.

Submitted for publication July 12, 2004. Accepted in revised form March 16, 2005.

